

23 April 1981

Tenth birthday for the environment

1982 will be a red-letter year in the history of the United Nations Environmental Programme, for then the newest of the United Nations agencies is planning to celebrate the tenth anniversary of its foundation, at the United Nations Conference on the Environment in Stockholm. The meeting of the agency's governing council planned to open on 13 May in Nairobi will be asked formally to approve the plans that have been laid to celebrate the occasion. There is, for example, to be a "session of a special character", a week-long occasion for mutual congratulation and exhortation, sandwiched into the normal proceedings in 1982. For this jamboree, nations which are not members of UNEP itself will be invited to attend; the only necessary qualification is that they should be members of some other United Nations agency or of the International Atomic Energy Agency. Those arriving in Nairobi in 1982 will be provided with a "60-70 page report" by the executive director, dealing with past and future environmental challenges, together with the annual document in which the agency spells out its estimate of the then "state of the environment". The meeting of the governing council next month will be asked to approve an outline of these documents and also a budget for the "session of a special character" which amounts to \$442,000, including \$163,000 by way of fees to the consultants who will be asked to prepare the special documents. It will be interesting to see how many of the delegates in May consider that, at something like \$2,500 a page, the cost may be excessive.

Compared with some other United Nations agencies, UNEP is neither ostentatiously extravagant nor obviously bigger than it needs to be. The agency rubs along on a mere \$5 million a year, spending perhaps twice as much on specific programmes to which member states contribute separately and voluntarily. Plainly it is not living off the hog, while it is not outrageous, at least by the standards set by other international organizations, that it should spend the best part of \$500,000 on a self-congratulatory conference a year from now. The more important question, which the delegates to Nairobi should ask themselves in May, is whether they should not spend their "session of a special character" the following year considering whether their agency is pointed in the right direction.

The most serious danger, clearly apparent in Stockholm in 1972, was that the agency might be allowed to tackle the variety of insoluble problems on which the founding fathers had set their hearts. That, fortunately, has been avoided, chiefly because governments were unwilling to provide the necessary funds. Since then, the agency has made its way in the world with a number of good works of a more unspectacular kind. It has helped to bring the convention on Pollution in the Mediterranean into being — and in the past few weeks has been seeking to do the same for the Caribbean (see p.622). It has competently taken over from the headquarters in New York the management of the United Nations committee on nuclear radiation.

Only in two respects can it be seriously faulted. Over the years, the agency has taken the view that a large part of its objective should be to "raise the consciousness" of member governments on environmental matters, which is entirely proper. But in doing so, the agency has taken to using the rhetoric of the extremists in the environmental movement. A simple booklet, published the other day and consisting of a reprint of the Stockholm Declaration and related documents, is loosely and unnecessarily called "In defence of the Earth". And, in attempting to draw attention to environmental problems, the agency has often fallen

into the trap of dramatizing them. A session of a special character in 1982 might usefully try to bring the agency back to earth in these respects. At the same time, UNEP could usefully be pointed in the direction of tasks that need to be carried out, and which are squarely within the terms of reference with which it was blessed at Stockholm. All member states, but developing countries in particular, are much in need of some way of ordering priorities in the commitment of resources to the improvement of the environment. UNEP could usefully shoulder these simple tasks, and earn itself a reputation in the process.

Going private

Under what circumstances can a research laboratory established under the wing of a protective government be legitimately turned into a private enterprise? Since its election nearly two years ago, the British government has been doing what it can to sell off assets which it has accumulated, sometimes by accident. A large block of shares in British Petroleum, the oil company, has already been disposed of, as well as 49 per cent of the shares in British Aerospace, a conglomerate of miscellaneous manufacturing concerns. In due course, the government also plans to sell off enterprises such as Cable and Wireless, the publicly owned international telecommunications concern. The chief objective of these sales of public assets is to raise cash that can be used to span the annual gap between government expenditure and revenue — and it is an open secret that the government is disappointed that "privatization" (as the jargon has it) has so far accomplished so little. Partly as a consequence, but also because the government is rightly persuaded that the strength of the civil service should be reduced, attention has now turned to the conversion of public laboratories into private enterprises of various kinds. The immediate danger is that the government may not be able to distinguish between the shedding of a laboratory and the shedding to the stock market of the public interest in a block of shares in some enterprise whose management is otherwise unchanged.

Several public laboratories are in various stages of being turned out into the private sector. The simplest case is that of the Radiochemical Centre at Amersham in Buckinghamshire, which supplies radioactive isotopes and chemicals containing them to an impressive array of customers throughout the world. Technically, the centre is constituted as a limited company, the shares in which are held by the UK Atomic Energy Authority, acting as a kind of public trustee. The 1,500 or so employees are not civil servants in the strict sense. The enterprise makes a tidy profit (nearly £1.5 million in 1979-1980), and has a marketable record of growth over the past decade. Although potential purchasers will be disappointed that the profit is merely 3.5 per cent of the turnover, there is every chance that this sound business could be sold as a going concern without interfering in the short run with the management of the laboratory. The government's chief concern, when it invites tenders for the centre a year or so from now, is likely simply to be that the Radiochemical Centre is not snapped up by some unwelcome bidder. Privatization will be found to have been painless.

The other laboratories on whose disposal the government has set its heart will not be as simply turned out into the cold. The chief of them is the Hydraulics Research Station, which has over several



years acquired an enviable reputation for model studies of hydraulic projects, many of them overseas (see *Nature* 9 April, p.438). Over the years, successive governments have expected different things of the laboratory. Sometimes the laboratory has acted as a hidden subsidy of the export trade. Now, with the doctrine that those who call the tune should pay the piper sweeping through Whitehall, it is natural enough that this laboratory should find itself one of the first on the edge of unfamiliar and apparently hostile territory. The laboratory will not, however, be the last to find itself in this position.

The most serious doubt about the government's intentions concerns the prospectus which the staff at the laboratory is being offered. The assumption seems to be that if the laboratory were established as a private company, it would be able to make its way in the world by selling its services to those who use them now. The most obvious difficulty is that the fees which have been charged to, say, construction companies looking for technical appraisals of proposed designs have not fully covered the overhead cost, and especially the cost of long-term research. Those working at the laboratory, and the Institute of Professional Civil Servants which represents the scientists among them, are naturally acutely aware of the risks they are being asked to run. The government seems so far to have failed to still their recalcitrance.

The underlying difficulty is that there is a genuine conflict of interest between the government and the staff at the laboratory. The government, in its role as employer, most probably adequately represents the national interest in holding that applied research laboratories providing a service for customers need not be a part of the public service. But those who work in these laboratories and who have become accustomed to the apparent but often illusory benefits of being civil servants cannot be expected spontaneously to subordinate their personal interests to the greater good. The most immediate fear, at the Hydraulics Research Station but also in other laboratories where privatization threatens, is that the private enterprise will not need, or will not be able to afford, all the staff now working for them. Another problem is that public servants, unlike employees in private concerns, know in advance what their financial prospects are. The government cannot hope without trouble to spin off this or any other laboratory without compensating those directly concerned for the risks they are expected to run. There should be a generous promise of support, over a period of some years, for the core of basic research essential to commercial success. To meet the proper fear of redundancy, the government should undertake that members of staff at the laboratory will not in, say, the next seven years be worse off than if they had remained in the public service. And, to sweeten the pill, arrangements should be made that, in the immediate future, members of the staff should actually be a little better off. For if the objective is to wean public servants to life in the commercial world, is it not in everybody's interest that they should be helped to appreciate the benefits of an equitable bargain?

Supersonic scandal

Ironically, the successful landing of the United States shuttle in the Mojave Desert last week coincided with the publication in Britain of a powerful economic case against the supersonic Concorde aircraft. At this stage, the shuttle is far from being an economic proposition. The cost of development (nearly \$10,000 million) is already substantial, it is unlikely that the National Aeronautics and Space Administration will seek to recover any substantial part of this in the charges made to the ultimate users of the spacecraft and, of course, unforeseen snags may add still further to the cost of the whole development programme. But the shuttle has promise, and probably a future. Concorde, by contrast, is a technological project with a past but no future. The House of Commons Industry and Trade Committee last week published figures showing that Concorde is at present uneconomic and likely to become even more of a drain on the

financial resources of its operator, British Airways, as fuel costs increase. When the British government is skimping on public expenditure in all its works, Concorde's capacity to keep flying must be an open and urgent question.

Since its inception in 1962, the Concorde project has been an assault on economic reality. The then Mr Peter Thorneycroft (now Lord Thorneycroft and chairman of the British Conservative Party), as a member of a British government seeking entry to the European Community, seems to have persuaded himself that an Anglo-French collaboration on the development of a supersonic aircraft would have political advantages outweighing the economic costs of the enterprise. In the event, Concorde did not overcome General de Gaulle's objections to British membership, but it was left to a later British government to discover that the Anglo-French agreement about Concorde was legally indissoluble. Technically, the promise of the aircraft was always limited. Concorde was conceived of as a first modest venture into the supersonic regime for a passenger aircraft. As an economic proposition, it was too small and too slow. Optimists, two decades ago, brushed aside these faults by saying that the first Concorde would quickly be followed by other economic aircraft. In the event, there have been no successors — or even serious suggestions that there should be. The British and French governments have between them spent £1,200 million on the development of this first aircraft, and yet must continue to subsidize their nationalized airlines to keep Concorde in the air.

The cost of keeping the aircraft in the air is still substantial. Both airlines have been forgiven the cost of developing the aircraft, and British Airways has also been forgiven the £160 million cost of buying six aircraft (one of which is now out of service). Yet even though the cost of a seat on Concorde is 20 per cent more than the cost of a first-class fare on a conventional aircraft travelling the same route, British Airways is able to come within sight of breaking even only on the route between London and New York. On the whole of its Concorde operations, the airline reckons that it will have to spend £65 million more than it earns from them in the next five years. The airline told the committee that some advantages accrue to it from being seen to have a continuing interest in supersonic flight, a calculation of the intangible that does not apparently prevent British Airways asking the British government for an annual subsidy for its Concorde operations. The committee concludes that the case for Concorde is so shaky that there should be commissioned an independent financial investigation of the continued operation of the aircraft. The surprise is merely that the committee did not recommend, on the evidence that it had gathered, that the airline should turn its back on Concorde before more money has been wasted on the enterprise. Why should taxpayers, or other airline users, continue to subsidize an enterprise without financial or technological promise?

The issue is deeper than that of how a nationalized airline manages its internal affairs. To the extent that supersonic aircraft, Concorde in particular, have become identified in the public mind with the process of innovation, the continued operation of a tiny fleet of loss-making aircraft can only injure the reputation of technology as a whole. The figures assembled by the Commons committee show that the average cost of carrying a passenger from London to New York by Concorde (£543) is more than three times the corresponding cost by conventional wide-bodied aircraft. Although a substantial part of the extra cost is that of fuel — in a sense, bad luck — the poor utilization of the aircraft (on average, 2.6 hours a day) and the high cost of persuading people to buy tickets on the machine also help to make Concorde operate in the red. These extra costs will hang like an albatross around the neck of British Airways for as long as it persists in flying an uneconomic supersonic aircraft across the Atlantic. For as long, British taxpayers will be asking what the British government can intend by exhorting everybody in sight to productive industrial innovation when it chooses to spend part of its income from taxation on an unproductive innovation such as Concorde. The enterprise, wrongly conceived twenty years ago, should now be brought speedily to an end.

Telecommunications monopoly eroded?

Profiting from leased lines made seemly

The most serious challenge yet to the British telecommunications monopoly came last week in a recipe for permitting outside access to the public network. A report by Professor Michael Beesley, *Liberalization of the use of British Telecommunications Network*, commissioned by the Department of Industry, argues that private persons and companies should in future be allowed to set up separate networks but also to market telecommunications services based on circuits rented from British Telecom.

In passing, the report provides a unique account of competition between telephone monopolies and those who use their circuits for marketing private services, notably in the United States. Thus MCI Inc., one of the survivors of the companies set up in the 1960s to provide data transmission services using lines leased from AT&T (the telephone company), told Beesley that 91 per cent of its business now stems from voice and not digital transmission.

In principle, the Beesley report is intended to allow the British government to fill in some gaps in the Telecommunications Bill, still being discussed in the House of Commons. When the bill was published last October, the Department of Industry promised that provision would be made for private interests to lease lines from British Telecom, reselling the services thus made possible. In practice, opposition from British Telecom — and the sheer unfamiliarity of the Beesley proposals on the British scene — may postpone these developments.

Jargon for these services is VANS, or "value added network services". Beesley concludes that in the domestic British market there should be no restriction on the use made of British Telecom circuits by those who rent them. British Telecom, the report says, should be required to publish the costs of renting circuits of different kinds, and the charges made should be unrelated to the proposed uses. (British Telecom's proposal that it should charge royalties on VANS is rejected.) British Telecom itself should be free to compete in fields such as data transmission. Ultimately, Beesley argues, liberalization of the monopoly should be extended to the international market.

The crux of the report is Beesley's rejection of the British Telecom protest that liberalization would allow private users to "skim the cream" off its

telecommunications market. British Telecom itself has estimated that unrestricted use of its domestic circuits would lead to revenue losses of between £30 million and £110 million, and that one consequence would be slower development of the telecommunications network in Britain. Beesley doubts the arithmetic (which implies a faster growth of VANS services in Britain than in the United States), says that the public monopoly could safeguard its position by adjusting tariffs more accurately to costs and argues that the partitioning of the telecommunications market is not, in any case, a "zero-sum game". The sale of VANS would increase the amount of business to be shared by British Telecom and those selling services based on leased circuits.

As yet, Beesley finds, there is no clear pattern of how British private users would exploit the freedom to use British Telecom circuits. The most obvious starting point would be the provision of long-distance telephone services, perhaps of lesser quality, more cheaply than British Telecom. But several organizations with existing private networks — one of them the nationalized industry British Rail — have ambitions to make more use of their existing telecommunications networks.

Although the Department of Industry, the department responsible for telecommunications, favours the principles underlying Beesley, the process of consultation is expected to take several months. Political and trade union opposition is likely to be fierce.

Postgraduate medical schools in danger

The University of London's postgraduate medical schools may be the first British academic institutions to bite the dust in the wake of government cutbacks in funds for higher education. A meeting between their representatives, Lord Annan (the vice-chancellor), the Secretary of State for Education and Science and the Health Minister, before Easter, did not secure more cash but did, apparently, convince the government of a problem that it had not known existed.

At risk are the 13 members of the British Postgraduate Medical Federation and two unassociated schools, the Royal Postgraduate Medical School and the London School of Hygiene and Tropical Medicine. Their problems began last October, when government support for new overseas students was cut. The decision that such students should pay "economic" fees has since been exacerbated by reductions of central funds for universities.

The members of the British federation, with 44 per cent of their students coming from overseas, decided to charge the minimum recommended fee of £5,000 even though their average costs are considerably more. A further problem is that a fair proportion of their students are salaried staff whose fees are waived, for which the schools now receive no offsetting compensation. The result is a substantial deficit which this academic year the federation may just be able to balance with help from central funds.

The fear now is that matters will get worse. The intake of overseas students is expected to fall considerably next October. Already several governments, in particular those of Malaysia and Hong Kong which traditionally send many medically qualified students to study in Britain, have said that in future they will be sending students to cheaper universities in other countries. If the expected shortfall

materializes, some of the federation's institutes may find themselves almost bankrupt by the end of the 1981–82 academic year.

That, according to Mr D. Innes Williams of the British federation, would have "disastrous consequences" for teaching and for London's health services. Most teaching and research would have to stop and the clinical staff would be absorbed into hospitals. One possibility, that could keep the institutes functioning although crippled, would be that the government should make up the costs after fees from overseas students have been deducted.

Another possibility, is that some of the smaller institutes could become departments of the major teaching hospitals. Discussions to incorporate the Institute of Dermatology into St Thomas's Hospital are already under way. The larger institutes, however, are hoping to manage on their own.

Although there are many similarities, each institute seems to have its own peculiar problems. The Institute of Basic Medical Sciences, for example, is considering merging with the Royal College of Surgeons, which would have to put up considerable funds to bail it out of its present difficulties.

The Institute of Cancer Research, which is affiliated to the university but receives no funds from it, is suffering not so much from the shortfall in overseas students as from a decision taken in 1977 by the Medical Research Council and the Cancer Research Campaign, which provide it with the bulk of its funds, to reduce their block grant. The institute is now in the position of having spent most of its reserves to make up the deficit and of having its block grant cut even further a year from now. Plans are already under way to persuade staff to retire early. Teaching may be cut in order to retain an adequate research base.

Judy Redfearn

Environmental asbestos

Dispute on levels

Brussels

The European Commission's proposed directives on asbestos have come under attack from all sides. Many of the criticisms put forward by the European Parliament and environmental groups relate to the tougher restrictions on crocidolite (blue asbestos) than those for chrysotile (white asbestos). The dispute highlights the present uncertainty as to what can be considered a "safe" exposure to asbestos fibres.

Following the adoption of a framework directive on asbestos in 1976, the Commission has been working on four implementing directives. Two have already been put forward: one on the marketing and use of certain dangerous substances and preparations; and the other dealing with the protection of workers from occupational exposure to asbestos.

The draft directive on marketing would ban crocidolite except in the manufacture of asbestos cement pipes, acid-resisting seals, gaskets and gland packings. The marketing and use of chrysotile and other asbestos fibres would be banned for thermal and acoustic insulation, air filtering and roadway surfacing unless "the harmful use of fibres is prevented".

The directive on occupational exposure has provoked the greatest discussion. This sets limits on the concentration of asbestos fibres in the air, together with a ban on asbestos spraying, access restrictions, the keeping of health records and a commitment to introducing suitable substitutes as soon as these are available.

The Commission's proposal assumes that crocidolite is five times more dangerous than other types of asbestos. The maximum concentration allowable for crocidolite is 0.2 fibres per millilitre of air, compared with 1 fibre per millilitre for all other types.

Evidence presented at a recent international symposium on the biological effects of mineral fibres supports the Commission's position. Although earlier research with animals had shown that fibres of the same length and diameter cause the same number of mesotheliomas in the lung, more recent research suggests that human lungs react differently to different types of asbestos. Chrysotile enters the human lung less easily than crocidolite and clears more rapidly.

Some members of the European Parliament's Environment and Public Health Committee feel the evidence is inconclusive, and others feel it is confusing to have two different limit values. The Commission thinks that if a single limit value were to be adopted, it should be at the lower level of 0.2 fibres until further research is carried out. But other committee members favour using the upper level of 1.0 fibres. **Jasper Becker**

Spinning off research

Further reductions in the scale of in-house research carried out within the British Department of Defence are promised in the annual "Statement of the Defence Estimates", published last week. In a brief section on research, the white paper says that, in the years ahead, further efforts will be made to place design and development contracts with industrial companies.

This policy, recommended in a study begun in 1979 under Lord Strathcona, Minister of State at the Department of Defence, also recommended that private industry should be given (on a contract basis) more responsibility for the provision of technical and other services to the British defence research establishments. The white paper records with pride a 15 per cent reduction of research manpower between 1974 and 1980, one third of this in the last two years of the period.

At the same time, the ministry is planning to switch resources within its research establishments to long-range studies of various kinds. It cites as one promising field of research a programme now under way to develop and anti-tank weapon in which a single projectile will release a clutch of separate missiles independently guided (with the help of microprocessors) towards separate tanks.

Space shuttle

Bonding problems

Washington

As the US space shuttle achieved its first orbital test flight last week, the National Aeronautics and Space Administration (NASA) was working on proposals for an alternative thermal protection system involving large carbon/carbon panels rather than the ceramic tiles used so far.

The tiling system has been the Achilles' heel of the whole shuttle programme. Difficulty in fixing the tiles securely to the aluminium skin of the orbiter was largely responsible for the two-year delay — and corresponding cost overruns — in the first launch, and the problems are still not solved.

During last week's flight, seventeen of the protective tiles covering the engine housing on the upper surface of the shuttle became partially or totally unstuck during the launch. NASA maintained optimism, saying that none of the more important tiles on the bottom of the shuttle had been damaged, and fortunately this optimism proved justified when the shuttle returned safely to Earth.

The problems that have plagued the development of the present heat protection system — which requires 31,000 tiles, each between six and eight inches square, to cover 70 per cent of the spacecraft's surface — have recently led the space agency to



Shuttle tiles, a sticky problem

look closely at different systems which might be used for future orbiters.

During re-entry, the wing tips and bottom surfaces of the shuttle are required to withstand temperatures of up to 1,370°C. In missiles and other conventional spacecraft, such temperatures have been tolerated by the use of an ablative heat shield which burns away during the heating; but as the space shuttle depends for its cost-effectiveness on reusability, a different solution had to be found.

Using what was the most advanced technology in the early 1970s, the shielding was designed to be provided by tiles of low-density high-purity silica fibre — sometimes referred to as "foamed glass" — made rigid by ceramic bonding. There are about 20,000 of these tiles on the bottom of the Columbia, the average size being about six inches square, and the thickness varying between 0.5 and 3.5 inches.

Each tile is bonded to a pad made of Nomex felt, and the total composite skin can heat up without placing stress on the silica, which is relatively easy to repair.

The ceramic tiles have the disadvantages, however, of being both brittle and not very strong. Furthermore, the aerodynamic stresses on the surface which tend to pull the tiles from the felt which binds them to the shuttle skin had been underestimated — whereas the bonding had originally been required to withstand pressures of up to seven pounds a square inch, in practice the stresses turned out to be twice as high.

The situation became particularly embarrassing when some temporary tiles were lost during the shuttle's transportation flight from California to the launch site in Florida two years ago. And subsequent weaknesses revealed by an intensive programme of "pull-testing" required a lengthy process of densification of many thousand tiles before NASA considered that they were safe.

In the light of these difficulties, NASA last year awarded a contract to Rockwell Corporation Inc. in Downey, California, for a study of alternative thermal protection systems. Three alternative systems

are being studied, each involving a combination of three different forms of protection. For temperatures less than 540°C, insulation would be provided by panels of titanium facing covering an insulating honeycomb. In the 540°C to 980°C range, the same arrangement would be used, this time using a superalloy known as Inconel 617.

The most significant change would be in the high temperature panels, where heat shielding would be achieved by panels of reinforced carbon/carbon, up to three feet square. Various insulating materials would be placed between the shielding and the aluminium surface. The panels would be mechanically attached to the shuttle, avoiding the bonding problem associated with the ceramic tiles. Carbon/carbon is already used in limited quantities for the nose-cap and the leading edges of the shuttle's wings.

But however the new systems perform, there is still the problem of cost to be resolved. Moving to a new system would be expensive, and NASA may not be able to afford a switch.

David Dickson

Nuclear reactor safety

Go for Super-Sara

Brussels

The months of deliberation in Brussels on the future of the Super-Sara project have ended with a decision to press ahead on the lines of the Commission's original proposal. The EEC's joint research centre at Ispra in Italy will now be able to start the second phase with funds totalling 50 million European Units of Account (£0.54 = 1 EUA).

The Super-Sara project will use the experimental ESSOR reactor to study loss of coolant and other accidents in water reactors. More than 200 staff have been left in limbo since last November when the member states first began to hesitate about releasing the bulk of the project's money. In March 1980, 3.31 million EUA was granted for feasibility studies after the project was agreed in principle. But when it came to making available the 40 million EUA for the major part of the project, Germany and others began to have cold feet. The negotiations over the past six months left Germany as the only member state withholding agreement. Germany has now given way to pressure from its partners but has insisted that the project be kept below its budgetary ceiling.

However, the European Commission's initial cost projections have had to be raised since they were first made last year. Automatic salary increases and rising equipment costs mean that 64 million EUA needs to be set aside to take the project up to 1983, when it is due to be reviewed.

It seems increasingly likely that it will overshoot its schedule by as much as two years.

Jasper Becker

Election in Israel

New deal for science?

Rehovot

If, as expected, Labour Party leader Shimon Peres becomes Israel's next Prime Minister after the election in June, he will be the first man with a good grasp of science and technology to hold that office. Although not a scientist or engineer himself, Peres was responsible for initiating many large-scale research and development projects undertaken by the Ministry of Defence, where he served from 1952 to 1977, becoming minister in 1974.

These projects contributed substantially to Israel's military capabilities and also laid the groundwork for its sophisticated aircraft industry (with a turnover in 1980 of \$500 million) and for its rapidly developing electronics industry. Indeed, a very large percentage of Israeli exports based on local research (an estimated \$1,000 million last year) were spin-offs from defence research set in motion by Peres.



Peres, Prime Minister in June?

Peres has placed heavy emphasis on science-related projects in outlining the goals of a Labour-led government. There are plans for a huge "science city" in the Galilee, heavy government subsidies for energy research, a continuation of the recently inaugurated Mediterranean-Dead Sea Canal project for hydrological power and, significantly, the establishment of two nuclear power stations in the Negev.

Israel already has two research reactors, a 5-MW facility supplied by the United States and set up just south of Tel Aviv, and a more controversial 24-MW reactor from France, located near the Negev town of Dimona. But, despite an absolute dependence on imported fossil fuel for the generation of electricity and the existence of the necessary nuclear technology, Israel still has no nuclear power stations.

This anomalous situation arises primarily because the United States, the most obvious supplier of such power stations, is prevented by the Anti-Proliferation Act from selling them until

Kamikaze to Halley

The European Space Agency (ESA) is hoping to enlist the help of the Soviet and Japanese space agencies in sending its Giotto spacecraft on a kamikaze mission to Halley's comet in 1986. ESA would like to send Giotto to within 200 km of the comet's nucleus to observe molecules evaporating from it before they recombine or are interfered with by the solar wind. Initial plans had assumed a closest approach of 1,000 km because of the difficulties of pinpointing the precise position of the comet in advance and because of the risk to the spacecraft of cometary dust. If Giotto gets within 200 km of the comet it may manage a couple of hours of observation before being destroyed.

Such a close approach can only be made, however, with help from the Russian space agency which is also sending a spacecraft to the comet. The hope is that data from the Russian craft, due to arrive at the comet four days before Giotto, could provide precise details of the comet's whereabouts and allow last-minute corrections to Giotto's course. The Russian spacecraft itself will not be able to approach nearer than a few thousand kilometres to the comet nucleus because, unlike Giotto, it lacks a dust shield.

The Japanese space agency will also be sending a spacecraft — at 135 kg the smallest of the three — to photograph the comet in the hydrogen Lyman alpha line. The three space agencies are discussing how they might exchange data gleaned from their separate missions, and next summer the three, together with the US National Aeronautics and Space Administration (NASA), will be meeting to discuss a collaborative project between all four agencies. Although NASA has no mission of its own to Halley's comet, it is organizing a "Halleywatch" which will make use of ground as well as space based observations.

Judy Redfearn

Israel signs the nuclear non-proliferation agreement and opens its nuclear facilities to inspection by the International Atomic Energy Agency.

It was only after Egypt signed this agreement in February that the United States agreed to provide that country with two nuclear power stations with a combined capacity of 2,000 MW. But Peres refuses to say whether a Labour-led government will do what all previous governments have refused to do, namely sign on the dotted line, or, if not, how he expects to get the reactors he proposes to place in the Negev.

Peres sees a close link between the country's economic development on the one hand and its scientific and technological development on the other. He accuses the present Begin government

of allowing the economy to stagnate and, as a result, setting in motion a brain drain that has deprived Israel of thousands of scientists and engineers.

But many specific questions remain unanswered. Will, for instance, Labour create a Science Authority or even a Ministry of Science to take over from the virtually moribund National Council for Research and Development?

Also unclear is the amount of government support that can be expected for basic research, hitherto dependent on overseas grants, and whether there will be more funds for universities, which have suffered a drop of more than 30 per cent in real income over the past few years.

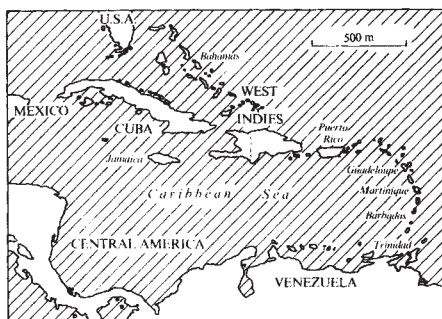
Peres will say no more than that he plans to "upgrade science, technology and education". How this will be reflected in Labour government budgets remains to be seen.

Nechemia Meyers

Caribbean anti-pollution plan Jamaican entente

Montego Bay, Jamaica

Temporarily forgetting their many political and cultural differences, government officials from 23 Caribbean nations met in Jamaica recently and agreed on guidelines for coordinating efforts to minimize the pollution of their seas and coastlines. The meeting was the latest in a series organized under the Regional Seas Programme of the United Nations Environment Programme (UNEP), most widely publicized of which have been those aimed at cleaning up the Mediterranean.



The plan of action agreed in Jamaica will probably have less immediate impact than the Mediterranean agreement, in part because the Caribbean is not a completely closed sea, and the level of pollution is therefore considerably lower.

Nevertheless, delegates to the Jamaican meeting stressed that the prime goal of the Caribbean plan should be preventative. According to Dr Stepan Keckes, director of the UNEP programme, "the governments have decided to tackle marine pollution not as a problem in itself but as the consequence of the problems which lie on the land".

The principal goal is to integrate concerns for the state of the marine environment into development plans for

the Caribbean region. The four projects already selected as being of top priority under the action programme include a regional oil-spill contingency plan, guidelines for managing watersheds, efforts to improve national environmental health services and broad measures to improve environmental education.

Relatively little money has been promised so far. A trust fund which is being set up as part of the action plan failed to reach its target of \$1.5 million, as a result of which UNEP secretary general Mostafa Tolba said that his organization would contribute an extra \$1.38 million. The principal donors to the fund have been Mexico, Venezuela and France, the first two being the main economic forces in the Caribbean region. Britain is prepared to make a contribution that reflects the specific interests of its dependent territories in the region (which include Antigua and the Caymans) but is unlikely to match the \$375,000 promised by France.

UNEP officials, however, as well as those from other aid bodies, hope that the plan will act as a stimulus for action at a national level. Jamaica in particular is hoping to take a leading role in coordinating development policies in the region. Addressing the closing session of the meeting, Jamaica's recently elected prime minister, Mr Edward Seaga, said that the agreement on marine pollution was "the greatest collaborative effort in the Caribbean to date, and may be a vanguard of future efforts to bond Caribbean nations in a common cause".

Jamaica is to become the base for a regional coordinating unit that will be established to facilitate the technical implementation of the action plan. The island has also put in a bid to become the home of the international sea-bed authority which would result from a successful outcome of the Law of the Sea negotiations now taking place at the United Nations in New York.

In the long term, the success of the Caribbean action plan is likely to depend primarily on procedures that individual states can be persuaded to adopt to minimize the sources of marine pollution. A top priority is to draw up a draft regional convention, and UNEP officials hope that a treaty committing countries to reducing sources of potentially damaging pollution can be agreed within two years.

The fate of the plan is also likely to be important to UNEP itself. Moves in Washington to reduce, or even eliminate, the US contribution to UNEP could cripple many of its programmes. US officials, however, are keen to cement political ties in the Caribbean — particularly with the new Jamaican government — and have indicated that the United States might in future be prepared to contribute to the trust fund. If UNEP can prove its value to the United States in the region, its medium-term prospects could be significantly improved.

David Dickson

Philosophy wrangle

Doubts persist about academic freedom in post-Tito Yugoslavia. Last month, the Yugoslav Secretariat for Information announced the reinstatement of the seven Belgrade professors of philosophy dismissed last January for their association during the early 1970s with the independent Marxist journal *Praxis*. Apparently the seven are to receive new appointments in an "Institute for Social Studies" specially created for them. Members of the group, however, say that no definite settlement has been reached.

Since *Praxis* was closed down in 1975, and the members of the group were suspended on 60 per cent salary, forbidden to teach or to publish within Yugoslavia, the authorities have repeatedly tried to persuade them to find new posts.

In June last year, the Serbian republic's law on universities was amended to permit dismissal after six months' suspension on the grounds of "damaging social interests". During the next six months, the authorities tried to persuade the suspended professors to find jobs. Two were even offered non-teaching posts within the philosophical faculty of Belgrade University, but they replied that they could not accept the post unless the other five were also reinstated.

Dismissal became effective in January, but the authorities later suggested that posts might be found in the Belgrade "Institute of Social Studies" — not part of the university but the authorities suggested that it could be rapidly converted into one.

Official plans, it turned out, envisaged that three of the group would retire, and the other four would be scattered among the existing centres of the Institute of Social Studies, covering areas such as "public opinion" and law. When these plans became known to the seven, they refused emphatically: they were, they said philosophers, and wished to work as such.

The vice-premier of Serbia, Milan Dragovic, who was acting as official negotiator, had already said that the transfer of the Institute for Social Studies to the university was nothing to do with the seven's demands, but was something that had long been necessary. He also conceded that a Centre of Philosophy within the institute would be a good idea.

On 3 April, the acting director of the institute, Dr Svetozar Culibrk, called in the whole group for talks, the first time that the authorities have been willing to negotiate with the professors *en bloc*. Dr Culibrk said that he "liked the idea" of a centre and would lay it before the institute's council. The seven are understood to be cautiously optimistic.

Vera Rich

CORRESPONDENCE

Irish evidence

SIR — Mr Peter H. Roberts's letter (*Nature* 19 March, p.184) offers the importation of infected Irish cattle as an explanation for the fluctuating incidence of bovine tuberculosis in the South West region between 1974 and 1980. It also says that the Zuckerman report carefully avoids this factor. I should like to set out the facts.

Table 10, on page 57 of the report, records the ministry's assessment of the sources of bovine tuberculosis infection in all infected cattle herds in Great Britain during 1972–1978, including imported Irish cattle. The figures for the South West and Sussex show 331 cases attributed to badgers and 4 (of which 2 were in Sussex) to imported Irish cattle. The totals for the rest of Great Britain are 0 and 171 respectively. It is little wonder therefore that Lord Zuckerman did not pursue the Irish cattle question when investigating infection in badgers in the South West.

The following breakdown of the figures for infection attributed to Irish cattle imports further detracts from the force of Mr Roberts's argument

	Great Britain	SW England
1972	59	1
1973	23	0
1974	22	0
1975	23	0
1976	23	1
1977	12	0
1978	13	0
TOTAL	175	2

Traditionally the Irish trade has been, and still is, mainly with the North-East of England and East Scotland. In the 2-year period before 1976 when, Mr Roberts suggests, cattle infected with bovine tuberculosis were imported from the Irish Republic, there was no increase in disease incidence from this source. What the figures do show is the effectiveness of the decision to introduce, in 1976, pre-export testing of Irish cattle.

W.H.G. REES

Ministry of Agriculture, Fisheries and Food,
Tolworth, Surrey, UK

Who is Nabi?

SIR — Readers may wish to know that the name of Isadore Nabi, the signatory of a recent letter criticizing my views on sociobiology and ethics (*Nature* 19 March, p.183) is fictitious. Should the writer ever make a statement over his own name, I hope he will confess that he lifted the two 1975 phrases of mine out of context in a way that reverses the meaning of one and makes it appear to contradict the other. I also trust that he will mention my later and fuller treatments of sociobiology and ethics in *On Human Nature* (1978) and *The Tanner Lectures on Human Values, Volume I* (1980).

EDWARD O. WILSON

Museum of Comparative Zoology,
Harvard University, Massachusetts, USA
Isadore Nabi is believed to be the pseudonym of Professor R.C. Lewontin of Harvard University — Editor, *Nature*.

Erratum

An incorrect spelling was given for the oak wilt fungus in the 26 March issue (*Nature* 290, p.284). The organism is *Ceratocystis fagacearum* and the beetle involved in its transfer is *Scolytus*.

Darwin's truths

SIR — The letter from your correspondents at the British Museum (Natural History) (*Nature* 12 March, p.82) relating to your editorial "Darwin's death at South Kensington" (26 February, p.735) fully supports your contention that something is amiss at that institution. The writers do not appear to understand the difference between a theory and a fact. When Darwin's *Origin of Species* was published in 1859, it was presented as a theory. It was then, and for long afterwards, proper to refer to it as "the theory of evolution". But since then the evidence for the theory has accumulated from many different sources and in many ways, among them natural and laboratory experiments. Wherever and by whatever means the theory has been tested, it has withstood every attempt at falsification.

The atom at the beginning of this century was a theory. No one today will doubt that it is a fact. Yet no one has ever seen an atom. Yet we have seen evolution in process before our very eyes in heritable changes in many forms, perhaps the most remarkable and obvious is the well-known case of industrial melanism in moths. What kind of proof do the South Kensington writers require before they will be willing to accept evolution as a fact? There are many differing theories concerning the mechanisms of evolution, and these are all to the good, but the fact of evolution as a process of change, surely, cannot be denied. The proofs for it are overwhelmingly clear.

ASHLEY MONTAGU

Department of Anthropology,
Princeton University,
New Jersey, USA

SIR — I am at a loss to understand what all the fuss is about concerning the phrase "if the theory of evolution is true". Darwin used it. I quote from *Origin of Species*:

"There is another and allied difficulty which is much more serious. I allude to the manner in which species belonging to several of the main divisions of the animal kingdom suddenly appear in the lowest fossiliferous rocks . . . If the theory of evolution be true it is indisputable that before the lowest Cambrian strata was deposited long periods elapsed as long or probably far longer than the whole interval from the Cambrian to the present day, and that during these periods the world swarmed with living creatures . . . The difficulty of assigning any good reason for the absence of vast piles of strata rich in fossils beneath the Cambrian system is very great."

This, coupled with the following quote by Dr W.R. Thompson, Fellow of the Royal Society, in the foreword of the *Origin of Species* (1956) does nothing to inspire confidence or belief in the current theories of how life started or progressed on Earth.

"It does appear to me in the first place that Darwin in the *Origin of Species* was not able to produce palaeontological evidence sufficient to prove his views, but that the evidence he did produce was adverse to them, and I may note that the position today is not notably different . . . As we know there is a great divergence of opinion among biologists not only about

the causes of evolution but even about the actual process, the divergence exists because the evidence is unsatisfactory and does not permit any certain conclusions." I am left wondering which theory requires the most blind faith.

GORDON SMITH

Liverpool, UK

Room for all

SIR — At first amused, I am now saddened by the "Death of Darwin" controversy (*Nature*, 26 February, p.735 *et seq.*). We have slipped back a hundred years: how long before letters signed Wilberforce and Huxley appear? Why "either-or"? Why cut off either right or left hands? Is there not room for Darwin, Hennig, God — and even Marx?

J.R. BAKER

Cambridge, UK

Popper's philosophy

SIR — Accusations that museum displays organized in the light of cladistic philosophy represent creeping Marxist-Leninism were bad enough. But when *Nature* sees fit to defend the scientific status of Darwinism on the grounds that "metaphysical theories are not necessarily bad theories"¹ then matters philosophical have truly gotten out of hand.

It should come as no surprise to those familiar with the writings of various cladists that the writings of Professor Popper on evolutionary theory are held in great esteem. After all, he has argued for years that no view of history, whether human, biological or technological, admits of the classification scientific. For Popper the events of history are unique and, thus, not amenable to systematic explanation via theories of any sort². Since cladists find the infusion of the least amount of theoretical insight into classification suspect in much the same way that their positivistically minded cousins, the phoneticists, did a decade or so ago, it is hardly surprising that they would greet the rediscovery of Professor Popper's writings from the 1940s and 1950s with great glee.

Popper's hostility to evolutionary theory explains the favour his views receive from systematists fond of cladism. But it does not explain why this same hostility should inform the judgments others make of the scientific status of Darwinism. Popper's views have been roundly and soundly criticized by numerous philosophers interested in the scientific status of evolutionary theory. The apparently timeless contentions that the theory is (1) tautologous, (2) unfalsifiable, (3) lacks predictive power and (4) lacks confirmation have shown to be false by a decade of scholarship in the philosophy of biology dating from the appearance of David Hull's *Philosophy of Biological Science*^{3,4}. Why *Nature* should choose the cladists' favourite philosophical authority over the ruminations of contemporary authors gives one pause not about politics but about the reading habits of the scientific community.

There are numerous reasons for not taking Professor Popper's criterion of demarcation of science and non-science seriously. Perhaps the most obvious is that it just does not cut well — cosmology and evolutionary theory wind up in the metaphysical hopper, astrology

and phrenology make the grade as science — false science but nevertheless science.

Numerous other contributors to the philosophy of science such as Nagel, Hempel, Hanson, Toulmin, Kuhn, Shapere, Lakatos, McMullin and many others have shown serious conceptual flaws in Popper's defence of the falsifiability criterion for demarcating science from non-science⁵.

Since the issue is so badly mauled in the *Nature* editorial let us take up the matter of the falsifiability of Darwinism since it seems of such great concern to cladists, creationists, and correspondents to this journal. The issue if stated as the question, "Is Darwinism falsifiable?" admits of a simple answer — yes.

If we are talking about the version of evolutionary theory propounded by Darwin in the *Origin*, then that theory has been shown false many times over. It claimed that all organic variation could be accounted for by natural selection and a tendency to inherit among all creatures. This is false. Mutation and recombination at the level of genes refute the adequacy of Darwin's posited mechanisms. Darwin claimed that the bulk of speciation occurred through anagenesis — evolution within a lineage. This is false. Cladogenesis or the splitting of an interbreeding population into two groups via natural barriers or other means accounts for the bulk of speciation. Darwin claimed that social behaviour exists in insects as a consequence of group benefit. This is false. Social behaviour exists among many insects as a consequence of kin selection, reciprocal altruism, parental manipulation or some combination of these mechanisms. This list could go on. The point is that evolutionary theory, like all theories in science is constantly tested, refined, modified, adapted and re-written. The issue confronting defenders of the validity and utility of evolutionary theory in its early or modern forms is not whether it is science or not — Darwin himself certainly settled that matter with admirable skill in the empirical and analogical evidence he brought forward in his own writings. Rather the question is how does one go about invalidating scientific theories that constantly change and evolve — an issue that ought to attract the attention of the cladists if not the creationists.

How true is the theory of evolution? It is as true a theory as there is in science. Which is to say it depends on who is asking and when. Of course the theory of evolution is not a fact. Evolution is a fact and the theory of evolution tries to explain this fact. But the theory of evolution surely passes muster as science, even on the out-dated grounds cited by Professor Popper in his early work. (Even he no longer believes in the adequacy of his early analyses of evolutionary theory⁷.) The real question that should concern scientists is whether they know enough about current thinking in the history and philosophy of science to know a sound theory when it stares them in the face — a circumstance unlikely to occur at the British Museum unless this ignorance is alleviated.

ARTHUR L. CAPLAN

The Hastings Center,
New York, USA

1. *Nature* 290, 75-76 (1981)
2. Popper, K. R. *The Poverty of Historicism* (Routledge & Kegan Paul, London, 1957).
3. David Hull, *The Philosophy of Biological Science*, (Prentice Hall, Englewood Cliffs, 1974).
4. Ruse, M. *Philosophy of Biology* (Hutchinson, London, 1973).
5. Suppe F. (ed.) *The Structure of Scientific Theories* 2nd edn (University of Illinois Press, Urbana, 1977).
6. Caplan, A. L. *Erkenntnis* 13, 261-278 (1978).
7. Ruse, M. *Philosophy Sci.* 44, 638-661 (1977).

Academia stagnant?

SIR — I see no reason for your concern (*Nature* 19 March, p.175) about the ability of British universities to recruit young scientists to academic posts. The present policy of recruiting young persons in their mid-twenties and sometimes without postdoctoral or international experience has so little merit that nothing would be lost if this recruitment of lecturers was totally ended. The best of our young scientists will wish to stay longer in full-time research and will continue to be available to the universities for years to come.

The rapid recruitment of academic staff in the 1960s expansion phase is unlikely to be the cause of stagnation in science departments. Where stagnation exists it is fairly certain that the department has defective leadership which might very well result from easy promotions gained in the years of expansion. Universities should, especially when resources are tight, recruit scientists into permanent teaching posts later in their careers than has been customary in Britain. A changed policy could provide continuing opportunities for the present generation of young scientists. Retirements and early retirements will improve some departments and at the same time allow new entrants or well justified internal promotions which ought to maintain the intellectual vitality of any department worth keeping.

J.R. PENSWICK

School of Biological Sciences,
University of East Anglia, Norwich, UK

Freedom from NATO

SIR — With the mounting concern in your columns on measures which scientists should take to ensure academic freedom for colleagues deprived of their rights, it is surprising that you publicize the NATO Advanced Study Institutes, that scientists are happy to perform on the stage which NATO sets, and that no mention of this paradox has been made in your correspondence.

No doubt these conferences fulfil a useful scientific function in promoting exchange of information in very congenial surroundings. Under normal circumstances, the argument that funds (from a rather dubious source) were being channelled to a good cause would have been sufficient to rationalize the situation. In some cases there is also the hope that the opinions of the scientific community will filter back along this channel, and that contact with sensitive institutions will lead to a more extensive dialogue and to a greater opportunity for diplomatic manoeuvres. Often, such occasions bring us into contact with scientists who have difficulty in meeting foreign workers because of governmental restrictions.

These conditions do not apply to attendance at NATO-supported conferences. There is little evidence that the meetings provide a forum for discussion with NATO officials in any way which would allow scientists to exercise their social responsibility. Nor could it be claimed that the activities of the conferences could not easily take place under other auspices. But most seriously, at a time when nuclear installations in Europe, products of scientific ingenuity, threaten the peace of our own and other countries, this close relationship between scientists and NATO can only be seen from outside as complete approval for (if not connivance with) the increasingly military orientation of our society.

This association with NATO must also affect the scientists themselves. After participation in any symposium it is difficult to leave without feeling some debt of gratitude to the organizers who lavish money and attention on their charges. Qualms about the ethics of the institution running it are henceforth stifled. Thus the NATO scientific conference is a highly effective public relations exercise both for the participants, and those who look to scientists for a lead in problems created by technology. How can Soviet colleagues be expected to take our views seriously when they see the strength of our opinions so cunningly controlled by political agencies? If our protests are to carry conviction in any field, we should dissociate ourselves from the influence of organizations like NATO, and take every opportunity to make this known publicly.

ROGER R.C. NEW

University of Liverpool, UK

Conservation sites

SIR — Many criticisms can be levelled at the system of Sites of Special Scientific Interest (SSSIs) but Muir's arguments (*Nature* 12 March p.82) are untenable.

Extinction may be "normal" but that is irrelevant to "value". To analogize, we will all die, but do not usually accept this as an excuse for murder. We are replacing complex ecosystems with much simpler landscapes, allowing for fewer species to exist or appear.

In a limited sense, rarity may support a value judgement of "biological deficiency", and the mere presence of a rarity may be poor grounds for conservation; but rare species can be good indicators of environmental quality when considered along with other evidence.

SSSI status may result in a "loss of freedom" of the owner. Losing such sites is a far greater "loss of freedom": it destroys our options and those of future generations, for we cannot replace them.

It is also misleading to allege just that protection of SSSIs is a cost to their owners: it prevents owners increasing their income by reclaiming the site, but that increase would itself be a cost to the taxpayer in terms of grants for reclamation, and later subsidies.

It is astonishing that someone from a university zoology department could contend that "Once . . . information is recorded and published . . . nothing new will be learned by preservation". No complete record of any site or habitat exists, and the same material can generate an infinite variety of "descriptions". Frequently we cannot turn to the literature for basic data to test new theories — we have to test them through fresh fieldwork designed specifically for that purpose. If we lose the field sites, ecology becomes metaphysics and not science, and to rely just on the environments we are creating today leaves us a severely restricted data base.

This necessity has fortunately been recognized in other fields. Once, excavators of bone caves destroyed vast storehouses of information by removing every vestige of material but failing to distinguish stratified deposits. Later workers recognized strata but again lost vital information by failing to extract pollen, rodent bones, and other small items. Now sections of deposit are left undisturbed for the future application of new ideas and techniques. Ecologists need similar opportunities.

KEVIN A. ROBERTS

Hoddesdon, Herts., UK

NEWS AND VIEWS

Genes of lymphocytes I: Diverse means to antibody diversity

from Miranda Robertson

AFTER fifteen years of controversy and five years of DNA cloning, immunologists are at last unanimous on how antibody diversity is generated. This remarkable eventuality is due to the accumulation over the past five years of evidence for almost every mechanism anyone has ever suggested; and to which has now been added yet another, of such baroque complexity that it could never have occurred to anyone to suggest it but for the evidence (albeit preliminary) in the DNA.

There used to be, broadly, three schools of thought, one of which adopted the extreme position that the entire antibody repertoire was encoded in a vast array of germ-line genes, and the other two of which favoured the somatic generation of variation on the basis of a limited number of germ-line genes. Two means were proposed for the somatic generation of variation. One was simple point mutation; but this on its own would not explain the distribution of the variation, which is concentrated in three "hypervariable" regions on the variable region gene (see Fig. 1). The other was recombination, represented in its most elegant and extreme form by the mini-gene theory of Kabat, who proposed that a range of genes coding for each of the three hyper-variable regions and the four less variable (framework) regions existed separately in the genome, to be brought together in variable combinations during the differentiation of the B lymphocyte. This theory however did not explain the evidence for point mutations in the framework regions. All these views have now been at least partly vindicated, as became clear at the massive immunology symposium in Paris last summer¹; and at a more recent and smaller symposium in Salt Lake City*, their proponents were still reeling with surprise at finding they no longer had anything to quarrel about.

Two things became clear very soon after the cloning and sequencing of the first immunoglobulin genes. First, there is quite a large number of germ-line genes: more than one for each of the known variable region subgroups (a subgroup is a family of

variable regions defined by similarity in the framework regions). Estimates have been made for mouse λ and κ light chain genes. Whereas the mouse λ V gene family contains only two genes², S. Cory (Walter and Eliza Hall Institute, Melbourne) now puts the size of the κ gene family at between 100 and 300; however, less is known about the possible numbers of heavy-chain variable region genes which she and her colleagues are in the process of investigating.

Second, the differentiation of a B lymphocyte is accompanied by the translocation of V region DNA. In the case of the light chain, one of the germ-line V genes, comprising the first three framework segments with two hypervariable regions and part of the third, is brought into proximity with one of a cluster of J genes comprising the rest of the third hypervariable region and the fourth framework segment. Since V genes may recombine with J genes at random, and with variation in the site of the join, this immediately generates considerable diversity. The scope for diversification is even greater in the case of the heavy-chain

genes, in which the V gene, which extends only to the end of the third framework region, recombines with a D gene comprising the third hypervariable region, and the D in turn recombines with one of a cluster of J genes coding for the fourth framework region. The D and J heavy-chain genes thus behave very much like the mini-genes postulated by Kabat; but Kabat never dreamed of the extraordinary properties of D genes that are beginning to emerge from the latest investigations of S. Tonegawa and his colleagues (Basel Institute of Immunology).

They have used a D region probe² to locate a cluster of eight homologous D genes in germ-line DNA. All eight genes are 17 base-pairs long; yet expressed D regions with substantial "core" homology to the germ-line genes range from thirteen to more than 40 base-pairs in length. This is too extreme to be explained by variations in the site of V-D-J recombination. Tonegawa has suggested that the explanation may lie in D-D joining, by a mechanism that would not result in simple tandem repetition of the homologous sequences. The mechanism depends on the

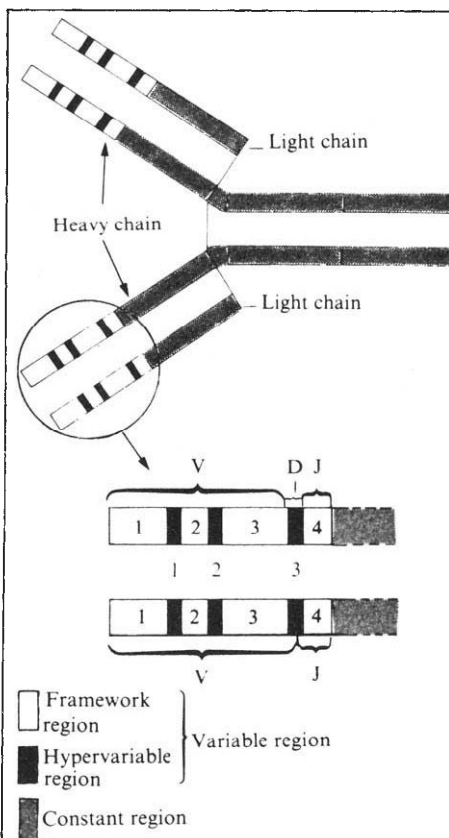


Fig 1. Immunoglobulin molecules are made up of four chains: one pair of heavy chains and one pair of light chains. Each chain is divided functionally into two regions: the variable region, which mediates antigen recognition, and the constant region, which in the case of the heavy chains mediates the effector function of the molecule.

The variable region can be further subdivided into four regions which vary only slightly from one antibody molecule to another, and are known as the framework regions, and three regions (the hypervariable regions) which are very much more variable and in the folded protein form the actual antigen-binding site.

Each light chain is now known to be coded by three separate genes: a C gene coding for the constant region, a J gene coding for the fourth framework region and part of the third hypervariable region, and a V gene coding for the first three framework regions including hypervariable regions one and two and part of hypervariable region three.

Each heavy chain is coded by four separate genes: a C gene coding for the constant region, a J gene coding for the fourth framework region, a D gene coding for the third hypervariable region, and a V gene coding for the first three framework regions including hypervariable regions one and two.

Miranda Robertson is associate editor of Nature.

*The ICN/UCLA Symposium on Immunoglobulin Idiotypes and their Expression was held in Salt Lake City, Utah, on February 8-15, 1981.

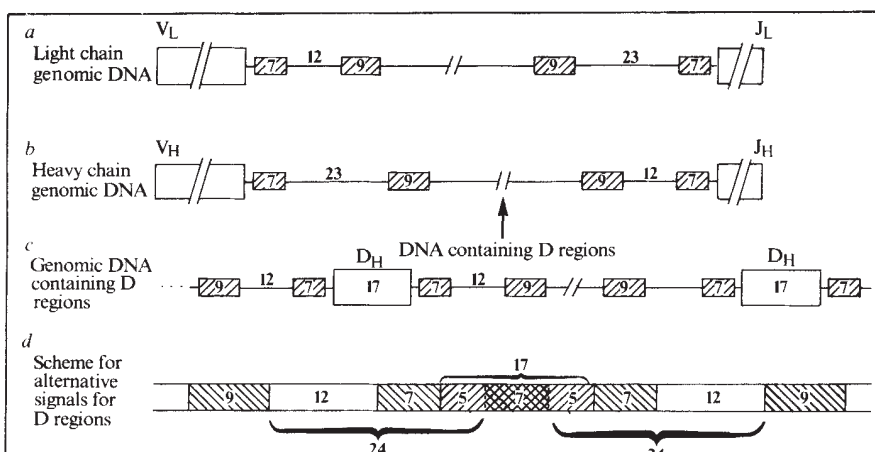


Fig. 2 The pattern of genomic sequences believed to mediate V-J and V-D-J joining in light (a) and heavy (b and c) chain variable region genes. Open rectangles are coding sequences, shaded rectangles are nonamer and heptamer signal sequences, and intervening lines are non-coding DNA. The figures 12 and 23 on the non-coding DNA refer to the length of the spacer regions between the heptamer and the nonamer, which are highly conserved sequences believed to act as recombination signals: the heptamer is followed by the nonamer after a space of 12 (strictly, 11 or 12) base pairs downstream of the V_L gene, and inversely, the nonamer is followed by the heptamer after a space of 23 (strictly, 22 or 24) base pairs upstream of the J_L gene (Fig. 2 a). Heavy-chain V genes are followed by the same signal sequences with the 23 spacer, and heavy-chain J genes are preceded by the inverse signal, again with a 23 spacer (Fig. 2 b). Early *et al.*⁴ and Sakano *et al.*⁵ surmised that recombination depends on the pairing of the signal sequences plus a 12-base-pair spacer with the inverse signal sequences plus a 23-base-pair spacer; they accordingly predicted that D_H genes should be flanked by signals with 12-base-pair spacers (Fig. 2 c). This has proved to be the case². If the 12/23 rule holds, however, it should make D-D joining impossible; and that is where the baroque complexity comes in.

Tonegawa and Kurosawa have proposed that each D region can treat the central 7 of its 17 base-pairs as the heptamer of the signal sequences and add the five coding base-pairs plus their adjoining heptamer, either upstream or downstream, to the adjoining spacer to make a 24 instead of a 12 base-pair spacer (Fig. 2 d). This would make it possible to add about 5 base-pairs of one D gene to another without violating the 12/23 (or 12/24) rule or producing tandem repeats, which are not found.

12/23 rule first enunciated by Early *et al.*³ for V-J recombination, and is described in detail in the legend to Fig. 2.

Even without D-D joining, the evidence for which is at best circumstantial, the recombination of a substantial repertoire of V genes at random with J and D genes would be sufficient to account for an antigen-binding repertoire well into the thousands. To this variation within chains must be added the variation generated by the association of different light chains with a given heavy chain. Since the expression of the heavy-chain genes precedes that of the light-chain genes, it has been assumed that each different heavy chain associates with a different light chain in the progeny of a single pre-B cell⁶. By fusing a pre-B cell line with a myeloma line that does not express L chains, W. M. Kuehl and his colleagues (Virginia University) have succeeded in inducing the expression of light chain genes of the pre-B cell and demonstrated that different rearrangements of the κ genes do in fact occur in the progeny⁷.

Finally, there is growing evidence for further diversification of the rearranged immunoglobulin genes by somatic mutation in the course of B cell ontogeny. In the absence of a B cell line whose ontogeny can be followed *in vitro*, the evidence is necessarily indirect. But the availability of cloned germ-line genes for

the V-regions of some antibodies has made it possible to compare the germ-line sequences with those of secreted immunoglobulins from B cells at different stages of maturation. At early stages of maturation, B cells synthesize immunoglobulin M; at later stages they synthesize immunoglobulins of other classes: IgG, IgA or IgE. Thus by comparing IgM and (say) IgG in turn with the homologous germ-line sequence, it is possible to detect any increase in departures from the germ-line gene that may have occurred as the cells switched from one class of immunoglobulin to the next.

Such increases have been reported both by A. Bothwell (Massachusetts Institute for Technology) and from the laboratory of L. Hood (California Institute of Technology), working with antibodies raised against nitrophenol (NP) and phosphorylcholine (PC) respectively. Antibodies raised against these antigens possess, in certain strains of mouse, the interesting and useful property that a very high proportion of the total immunoglobulin bears a characteristic antigenic determinant, or idiotype, of its own (NP^b in the case of anti-NP antibodies, and T15 in the case of anti-PC antibodies). But not all even of those antibodies bearing the same idiotype are by any means identical, and in the light of recent analyses this is hardly surprising. Both Bothwell¹⁷ and

Hood report that there is a cluster of several germ-line V genes all with substantial homology to the idiotype-bearing anti-NP and anti-PC proteins. The existence of such homologous V-gene clusters (as well as Tonegawa's homologous D gene clusters) strongly suggests duplication of genes during evolution — particularly since two of the five homologous genes sequenced by Hood and his collaborators are actually identical.

As it turns out, comparison of the partial amino-acid sequences of several T15 proteins with the germ-line sequences shows that all the proteins are derived from the two identical germ-line genes, but with numerous substitutions, notably in the proteins of ontogenetically later classes (IgG and IgA)⁸. None of these substitutions can be explained except by mutation, with one exception which could (but need not) be due to recombination between two of the germ-line V genes.

The sum of all these recent data leaves us with an embarrassing riches of diversity-generating mechanisms, but some important outstanding questions about their deployment. The first and most obvious is exactly how many germ-line genes there are. It is believed that there are four functional J genes each for the light and heavy chains; but estimates of the numbers of V genes still range between 500 and 1,000, and little is known of the total size of the D-gene family, of which that sequenced by Tonegawa and his associates may represent only a branch.

The second and more challenging is whether a special mechanism, still undiscovered, is needed to explain the concentration of amino-acid substitutions in the hypervariable regions. Both germ-line sequence and somatic recombination go some way towards explaining this pattern of variation. Related germ-line genes differ from one another more in the hypervariable regions than in framework-coding regions; and variable V-J and V-D-J joining account for further variation in the third hypervariable region — particularly in the case of V-D-J joining. But Gearhart *et al.*⁸ have shown that somatic mutations too are clustered, at hypervariable regions 1 and 2.

This could in principle be explained without recourse to further special mechanisms, simply by somatic selection. B cells whose surface immunoglobulin does not bind antigen produce no progeny: thus those expressing mutations in the hypervariable regions which form the antigen-binding site will tend to be selected. Selection will also occasionally operate on mutations affecting framework sequences: although it is the hypervariable regions that actually make contact with antigen, the framework regions may influence the manner in which they do so — for example, framework variants of an immunoglobulin which in its original form bound PC both on its own and complexed to a carrier will no longer bind the PC-

carrier complex (M.D. Scharff, Albert Einstein College of Medicine). Both the clustered substitutions in the hypervariable regions, and the occasional ones in the framework can therefore be accounted for by selection for antigen binding.

It would be a mistake however to assume that all mutations in expressed immunoglobulin genes must have been specifically selected: indeed, some are silent and could never have been subjected to selection. These are presumably the "neutral mutations" of the immune system. But if the pattern of mutations requires no special explanation, it is possible that the rate of mutation may. Certainly many immunologists are wondering about possible mutagenic reactions associated with the class switch. Notwithstanding the considerable difficulties associated with the elucidation of molecular mechanisms of immunoglobulin gene rearrangement,

the issue of mutation rate may be resolved more quickly by research at this level than by arguments about sequences, because arguments about the mutation and selection of sequences depend on a detailed understanding of the relationship between the sequence of an immunoglobulin and its antigenbinding properties and this is usually unknown.

However, each antibody molecule can now be seen as the end product of a series of selective steps, beginning with the germ-line genes. The germ-line V, J and D genes have presumably been selected in the course of evolution for their ability to produce proteins that will bind to common environmental antigens. There follow at least two rounds of somatic selection by antigen. The first acts on the V-J and V-D-J recombinations that produce IgM. It is followed by at least one more round of somatic selection favouring mutations that produce efficient binding to the same antigen, or indeed to any other antigen present in large enough quantities to encounter the mutant cell. A detailed analysis of hybridoma proteins from immunized and hyperimmunized mice, such as Bothwell and his colleagues have already begun, may help to show how far antigenic selection could account for the range of antibodies a single mouse is able to produce against a single antigen.

coupled with the missing oceanic crust, as arising from Earth expansion during the late Proterozoic. Pure Earth expansion has the effect of leaving continental centres in the same coordinate position and so has no effect on the polar wandering paths. The problem is to see how the relative continental motions during the Phanerozoic, which palaeomagnetism brings out so clearly, could have returned the continents to the relative positions they had in the Proterozoic.

The main counter-arguments to expansion were the well known length of day calculation (Runcorn *Nature* 204; 823, 1964) based on daily and annual growth increments in corals, the lack of secular change in global sea level over the Phanerozoic (Vail *et al. Mem. Am. Ass. petrol. Geol.* 26; 83, 1977) coupled with near stability in the volume of the hydrosphere (Anderson *Rev. Geophys. Space Phys.* 13; 37, 1975), and my demonstration that for the past 5 years, half or more of the subduction required by plate tectonics on an Earth of fixed radius is actually occurring. These arguments give a maximum rate of change in Earth radius of 0.1 cm yr^{-1} since the Devonian, in gross conflict with the idea of accelerating expansion during the late Phanerozoic. This conclusion is backed up by the provinciality of Palaeozoic faunas, pointed out by Clive Burrett, which suggests that major ocean basins existed at that time.

All this, however, fails to provide any hard evidence about palaeoradius in the Proterozoic, to which attention is now focused. Only Stewart's (*J. geol. Soc. Lond.* 133; 281, 1977) palaeogravity method, which compares computed geotherms with mineralogical evidence for the actual $P-T$ conditions in the crust and upper mantle, gives any help here. Palaeoradius can easily be calculated from the surface gravity field provided that Earth mass and the universal constant of gravitation are known. Gravity seems to have remained within a factor of two of its present value throughout Earth history, implying a rate of change in radius $< 0.1 \text{ cm yr}^{-1}$ and hence limiting late Proterozoic expansion to $< 1,000 \text{ km}$. This is too small by a factor of 2.5 to solve the problem of the missing oceanic crust.

It is to be hoped that the basically sound palaeomagnetic technique for measuring palaeoradius will soon be applied to early Proterozoic basic dykes across a stable continent such as Canada. Using samples collected specifically for the purpose, it should be possible to reach a definite decision about the likelihood of late Proterozoic Earth expansion. Published palaeoradius estimates using the palaeomagnetic technique have always used samples from Phanerozoic rocks and even these, it was pointed out at the meeting, did not always come from stable continental areas as required by theory, so that the validity of the palaeoradii was open to doubt. □

1. see Robertson, M. *Nature* 287, 390 (1980).
2. Brack, C., Hiram, M., Lenhard-Schuller, R. & Tonegawa, S. *Cell* 15, 1 (1978).
3. Kurosawa, Y., von Boehm, H., Haas, W., Sahano, H., Traunecker, R. & Tonegawa, S. *Nature* 290, 565 (1981).
4. Early, P. *et al. Cell* 19, 981 (1980).
5. Sakano, H. *et al. Nature* 286, 676 (1980).
6. Burrows, P., Lejeune, M. & Kearney, J.F. *Nature* 280, 838 (1979).
7. Riley, S.C., Brock, E.J. & Kuehl, W.M. *Nature* 289, 804 (1981).
8. Gearhart, P., Johnson, N.D., Douglas, R. & Hood, L. *Nature* in the press.
9. Bothwell, A.L.M. *et al. Cell* in the press.

The expanding Earth

from A.D. Stewart

TWENTY-FIVE years ago the University of Tasmania hosted a conference on continental drift organized by Professor S.W. Carey, at which he presented his now famous review documenting the detailed geological evidence for continental break-up and separation during the Mesozoic and Cenozoic, with explanations for major Earth structures like oroclines (hair-pin bends in mountain belts) and sphenochasms (wedge-shaped oceanic areas like the Bay of Biscay) thrown in for good measure. Most of Carey's views on inter- and intracontinental motions are now, of course, generally accepted but the basic mechanism he postulated — Earth expansion sufficient to have generated all ocean basins since the Palaeozoic — engenders in most geologists the same feelings of incredulity that continental drift did 25 years ago.

February 1981 saw the sequel to the Tasmanian meeting — the Expanding Earth Symposium at the University of Sydney, again organized by Carey — at which two new lines of evidence supporting major Earth expansion were discussed. The first was the work of Andy Glikson

(*Geology* 7; 449, 1979) which highlights the lack of geochemical evidence for oceanic crust, or indeed any mantle-derived crust, surviving from the Proterozoic, and the palaeomagnetic and stratigraphical evidence for continental integrity over this period. This was backed up by Keith Crook's analysis of the global abundance of ophiolites and turbidites, both generally accepted indicators of oceanic crust. Their abundance decreases exponentially with increasing age back to the end of the Proterozoic ($\sim 10^9$ years ago) when they disappear completely from the geological record. This is not just a sampling effect because turbidite sediments are well known from the Archaean ($> 2.5 \times 10^9$ yr). Crook's interpretation of these data is similar to Glikson's — there were no ocean basins in the Proterozoic because the Earth had only about 60 per cent of its present radius.

The second new line of evidence favouring expansion comes from palaeomagnetism. Embleton and co-workers (*J. Geophys.* 49; 20, 1981) showed that for quite long periods during the Proterozoic, continental blocks such as North America and Australia in their present relative positions had common polar wandering paths. They interpret this surprising result,

A.D. Stewart is in the Department of Geology at Reading University.

Plesiosaurs — rigging and ballasting

from Michael A. Taylor

ALTHOUGH plesiosaurs were among the dominant marine reptiles of the Mesozoic era, it was difficult to see how they hunted their prey for they had large air-filled lungs and would have been too buoyant to dive underwater easily. In a recent paper D.G. Darby and R.W. Ojakangas (*J. Paleont.* **54**, 548; 1980) suggest that plesiosaurs neutralised their upward buoyancy by swallowing masses of pebbles and retaining them in their stomachs in much the same way as modern crocodiles do.

At one time it was thought that the stomach stones of crocodiles and plesiosaurs helped in grinding food in the stomach. Darby and Ojakangas draw attention to the work of H.B. Cott (*Zool. Soc. London Trans.* **29**, 211; 1961) who found fragile but unpulverised bones and shells of prey along with stones in crocodile stomachs. He went on to show that crocodiles swallow stones to use as ballast so they can lie hidden under the river surface with only the eyes and nostrils protruding above water, until some prey animal comes close enough to be taken. Crocodiles can also lie on the river bottom lodged against the current. The mass of pebbles lying forward in the belly stabilises the animal in roll about the longitudinal axis, like the keel of a boat, and corrects its tailheaviness allowing it to lie horizontal in the water. The weight of the pebbles also helps the crocodile to drag large prey underwater so that they can be drowned. Darby and Ojakangas point out that stones and unpulverised prey have also been seen in a fossil plesiosaur (B Brown *Science* **20**, 184; 1904) and so conclude that plesiosaurs also used their stomach stones for ballast.

Their conclusion has implications for our ideas of the hunting strategies of plesiosaurs. If they neutralised virtually all their upward buoyancy, they could stay dived without using much limb action, in contrast to the extinct ichthyosaurs and modern whales which have no stomach stones, but are always on the move when looking for prey and can produce a compensating downthrust with their fins. One group of plesiosaurs, the plesiosauroids, had small heads on exceedingly long necks which gave the animals a most unstreamlined shape and perhaps made them too slow to pursue prey in open water. One wonders if they lurked in ambush, or searched slowly over the bottom ready to strike out their necks at any passing prey. The other main group of plesiosaurs, the pliosauroids, were large carnivores perhaps comparable with the modern killer and sperm whales. They may

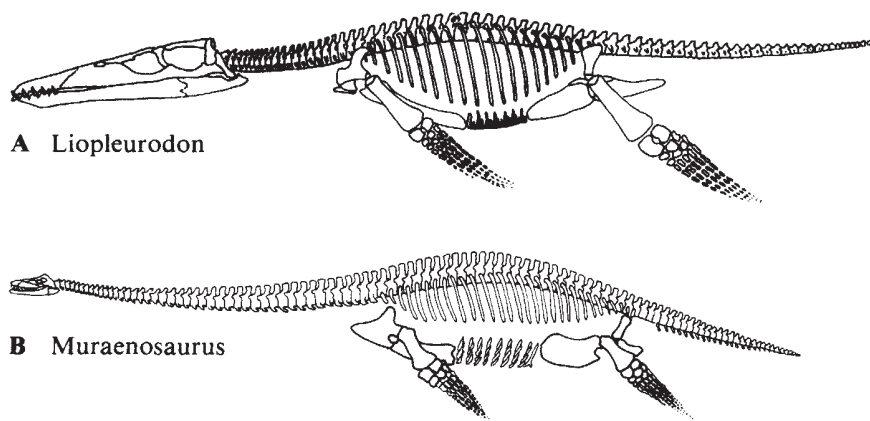


Fig.1 Representatives of the two main groups of plesiosaurs: A, the pliosauroid *Liopleurodon* (~11m); B, the plesiosauroid *Muraenosaurus* (~5m) (after Robinson 1975).

have hid unmoving till prey came near enough to be run down in a short sprint, but they were quite streamlined with their large heads and short necks and might equally have been open-water cruising predators. Confirmation of these hypotheses must wait for an analysis of the functional anatomy of the head and neck of these forms (see Fig. 1).

Darby and Ojakangas also recovered stomach stones from a plesiosaur found in the Upper Cretaceous of Montana and concluded from an analysis of the shapes of the pebbles that they were collected in an estuary rather than from a shingle beach. It does indeed seem likely that plesiosaurs would forage in the sheltered and productive tidal waters of an estuary rather than off an exposed and comparatively foodless beach. The plesiosaur may, however, have tended to select certain shapes of pebble, so that its stomach stones may not necessarily be a reliable guide to their source.

The structure and function of the distinctively domed body, sprouting four large broad-bladed limbs, has itself been

the subject of an elegant analysis in recent years (*N. Jb. Geol. Paläont. Abh.* **149**, 286; 1975 and **153**, 86; 1977).

Jane Ann Robinson distinguishes between rowing and underwater 'flying' as the two possible modes of action of the limbs. These two modes differ radically in limb form and action. In the power stroke of a rowing action, the limb sweeps backwards, while its vertically aligned blade bites into the water and levers the animal forwards. The limb then returns forward to start another power stroke. A limb which is adapted to rowing is therefore rigid and oar-shaped, broadening at the end, and moves in the horizontal plane. The limb of a plesiosaur is of quite different design and has the characteristic adaptations to underwater flight. The limb is flat with an aerofoil (or hydrofoil) cross-section so that when it moves through the water with the plane of the limb edge-on to the direction of motion through the water it generates a lift force. This lift force is perpendicular to the plane of the limb and its direction of motion through the water. The plesiosaur limb has therefore to move in the vertical plane to produce the horizontally and forwardly directed propulsive force, unlike a rowing limb which moves in the horizontal plane (see Fig. 2). The plesiosaur limb tapers towards its tip to reduce drag, unlike the oar-like shape of the rowing limb. The plesiosaur limb has just enough flexibility to allow trimming and to damp out drag-producing eddies, which is, again, quite unlike the rowing limb which is rigid and relies on its production of drag to bite into the water and thus gain leverage and efficacy. Indeed, the limbs of plesiosaurs share these features with the forelimbs of marine turtles, sealions and penguins, which all act as hydrofoils. Robinson has therefore been able to base her reconstruction of plesiosaur limb action

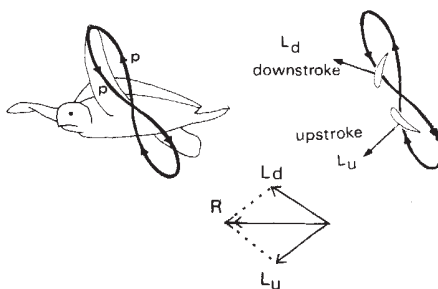


Fig. 2 Plesiosaur limbs worked like those of turtles, shown here. The hydrofoil surfaces swept up and down producing lift forces L_u and L_d which averaged out over the cycle to a forwardly directed propulsive resultant force R as shown in the parallelogram of forces (after Robinson 1975).

Michael A. Taylor is in the University Museum, Oxford.

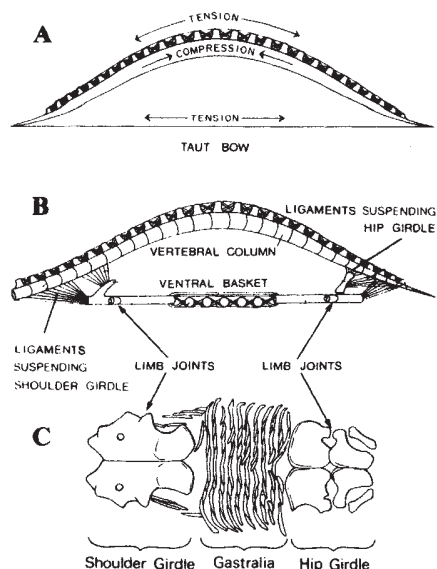


Fig. 3 The shoulder and hip girdles were suspended within the ventral basket which was held under the vertebral column like the string in an archer's bow. This elastic suspension could absorb the up-and-down motion of the limb girdles at each beat of the limbs, while transmitting the propulsive force to the body. A and B, side views of the body; C is a plan view of the ventral basket (after Robinson 1977).

partly on the action of the forelimbs in these modern animals.

The distinctive shoulder and hip girdles of plesiosaurs were broad, flat bony plates with limb joints on opposite edges. The main propulsive muscle masses lay on top and under the girdles, and their tendons attached to the dorsal and ventral surfaces of the limbs and swept them powerfully upwards and downwards. Supplementary muscles controlled the trim of the wings so that on both up- and down-stroke the hydrofoil surfaces of the limbs, moving in the vertical plane, generated a resultant 'time-averaged' lift force forwards. Effectively, therefore, each pair of limbs mounted on its girdle was a self-contained power unit producing a forwardly directed propulsive force.

How were the two 'power modules' supported within the domed body, so as to transmit the propulsive forces? Plesiosaurs had a distinctive 'ventral basket' made up from the two limb girdles bound together by ligaments running onto and between the gastralia, the splint-like bones of the middle abdominal floor. Robinson shows that this acted as a string to the 'archer's bow' formed by the vertebral column. The contraction of the muscles between the dorsal processes of the vertebrae bent the ends of the column upwards and pulled taut the ventral basket suspended from the vertebral column by its ligamentous attachments. With the ribs, this formed a remarkable suspension system restraining the girdles while allowing some elasticity to absorb the up-and-down motion of the girdles with each beat of the limbs (see Fig. 3). The ability to vary independently the

action of each limb would be most useful, as it made possible rapid changes of direction in pursuit of prey, or when diving from a position at the surface. The propulsive forces of the four limbs would then be asymmetric in magnitude and direction, but they could still be sustained by the girdle support system. One can also conclude that the plesiosaurs could move and breed on land, *pace* Robinson, as the arched vertebral column prevented the collapse of the lungs. Ichthyosaurs and modern whales cannot move on land and perform give birth in the water to live young; certainly, there are specimens of ichthyosaurs — but not plesiosaurs — with unborn young still in the body cavity. Plesiosaurs may have given birth to live young or laid eggs in pits scraped out with their tails, like modern turtles.

The plesiosaur tail was not very long and part of it was taken up in the 'archer's bow' system. The tail vertebrae bear no evidence of the attachment of a substantial propulsive tailfin — which would not be needed, with four powerful limbs — but plesiosaurs did have small vertical tailfins, presumably used for stabilising the animal directionally and for controlling direction of motion.

The limb and body plan of plesiosaurs, so uniform within the group, is radically different in form and function from the original terrestrial reptilian plan, which was adapted to support the body against the vertical pull of gravity while walking on land. The nothosaurs of the Triassic were small marine carnivores not very different in body or limbs from the terrestrial reptiles. However, the rare nothosaur *Pistosaurus* is morphologically intermediate between other nothosaurs and plesiosaurs and is presumably closely related to the ancestors of plesiosaurs. Nobody has as yet done an analysis of these animals comparable to Robinson's work and so we do not know when the characteristic locomotor pattern of plesiosaurs actually evolved. The plesiosaurs diverged rapidly into their variously adapted lineages at the very start of the Jurassic, in a good example of the macroevolutionary concept of 'adaptive radiation'. Did this event coincide with, and was it caused by the full development of the plesiosaur locomotor plan, or was there some other factor, such as the mass extinction of potential competitors, which allowed the plesiosaurs to exploit the potential of a previously evolved locomotor plan? □

Strategies of viral oncogenesis

from John Wyke

STUDIES on the molecular mechanisms of neoplasia by oncogenic RNA viruses (retroviruses) have in the past concentrated on agents that transform cells in tissue culture. These viruses contain genetic information, oncogenes, that seem essential for the induction and/or maintenance of cell transformation and that are distinct from the genes required for virus replication. Viral oncogenes are believed to be derived from the homologous sequences, present in normal vertebrate cells, that were presumably acquired by the viruses during their evolution. Their reintroduction into the cell under the control of viral transcriptional promoters leads to aberrant oncogene expression and resultant cell neoplasia.

Attention has recently turned to the many retroviruses that do not transform cells in available tissue culture systems but nonetheless induce tumours in animals. These viruses seem to possess only those genes needed for their replication. Without specific oncogenes, how do they induce neoplasia? There are, *a priori*, several possibilities: (1) expression of viral replicative genes themselves may directly or indirectly alter cell growth; (2) since the DNA copy of the retroviral genome, the

provirus, integrates into the cell DNA, it may act as an insertional mutagen, its presence disrupting normal cell gene expression; or (3) the inserted provirus may subvert functions more specifically, by placing particular cell genes under the influence of viral promoters. Two recent papers^{1,2} provide evidence for the last possibility from the avian lymphoid leukaemia viruses (ALV).

Both studies have taken a similar approach, based on recent understanding of the nuances of provirus structure and function that are summarized in the diagram. The coding sequences in the viral RNA (*gag-pol-env* in diagram) are flanked by 5' and 3' unique segments, U5 and U3, and a short terminal redundancy, R. These are thought to be involved in synthesis of viral DNA which eventually becomes integrated as a linear provirus (shown in diagram). The proviral DNA is longer than the parent RNA, U3 being repeated at the 5' end with U5 duplicated at the 3' end of the molecule. The viral genes are thus bounded at each end of the provirus by a long terminal repeat (LTR) sequence, comprising the regions U3-R-U5. Sequencing and *in vitro* transcription studies indicate that U3 contains a putative promoter sequence, and the promoter in the left-hand LTR is thought to direct transcription of progeny viral RNA. Because of the redundancy of the LTR

John Wyke is in the Tumour Virology Laboratory, Imperial Cancer Research Fund Laboratories, London.

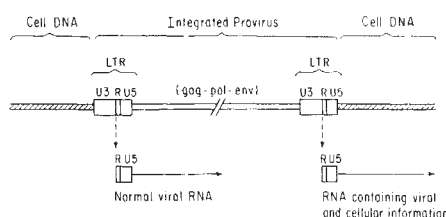
there is also a promoter at the right-hand end of the integrated provirus that seems to have no role in the viral life cycle. However, preliminary evidence indicated that it might also be active and occasionally direct transcription into neighbouring host DNA sequences³.

Starting from this point, Payne *et al.*¹ and Neel *et al.*² investigated integrated provirus structure and function in a number of advanced lymphomas induced in chickens by several different ALV. They drew the following conclusions. (1) The cells from individual tumours showed provirus integration at identifiable sites. This suggested that the tumours arose largely, if not entirely, from a single infected cell, implying that tumour induction is a rare event. Although many tumours contained several proviruses, those which carried only a single provirus were most useful for further interpretation. (2) Tumours from different birds often showed similarities in the sites of provirus integration. This evidence for tumour-specific integration sites was not strong but was significant in view of the many previous studies showing that proviruses are able to integrate at numerous, perhaps even random, locations. (3) Many tumours contained incomplete proviruses even when only one provirus was present (these latter cells did not contain normal viral RNA species). Thus viral replicative functions are not needed for the maintenance of neoplasia; indeed, immune mechanisms directed against viral antigens might favour the survival of cells whose proviruses do not direct the synthesis of viral proteins. (4) All tumours contained at least one LTR. (5) All tumours contained virus-related RNAs that possessed too little virus-specific material to be revealed by a probe representing the whole provirus but could be detected with a probe for the R-U5 sequences. These RNAs, which fall into only two or three size classes, were expressed at a high level and are thought to represent transcription initiated in a viral LTR (and hence containing U5 sequences) and continued into downstream host sequences (see diagram).

On the basis of these findings Payne *et al.* and Neel *et al.* propose that ALV cause tumours by inserting a viral LTR, containing a promoter, adjacent to cell sequences that are then expressed under viral control. The evidence for proviral integration in a specific region of the cell DNA and the limited number of size classes of RNA found among virus-related transcripts suggested that viral promoter insertion might lead to neoplasia only if it occurred near one or a few appropriate cell genes.

This possibility was pursued by Hayward *et al.* (see *Nature* April 9, 290; 475, 1981), who reasoned that ALV might 'switch on' the cell homologues of the known oncogenes of transforming retroviruses. Accordingly, they screened lymphoma

RNA with cDNA probes specific for five different viral oncogenes. In general, there was no enhanced expression of sequences homologous to four of these, but in 12 out of 14 tumours Hayward *et al.* found increased expression of RNA homologous to the oncogene of the acute leukaemia virus MC29 (an oncogene variously called *mac* or *myc*). The *myc*-specific RNA, which was not expressed above background levels in non-tumour tissue from



Structure and transcriptional products of the integrated ALV provirus.

the same diseased birds, co-migrated with the same limited size classes of RNA that are detected in tumours with a probe for R-U5 viral sequences, suggesting that the *myc* sequences in the RNA were covalently linked to viral U5 sequences. This conclusion was strengthened by restriction enzyme digests of tumour DNA which showed that fragments containing proviral and flanking cell DNA were detectable with *myc*-specific probes, indicating that the provirus is integrated near the cell homologue of *myc* (*c-myc*). They concluded that in the bulk of ALV-induced lymphomas a viral LTR integrates near *c-myc*, enhances its expression, and this leads to neoplasia.

The data in favour of this promoter insertion mechanism of oncogenesis are attractive but there are a number of inconsistencies which must be clarified before this concept can be accepted. (1) In a minority of lymphomas there is no evidence for enhanced *c-myc* expression. (2) MC29 causes carcinomas and myelocytomas, not the B-cell lymphomas associated with ALV infection. In MC29 *myc* is expressed as a fusion protein, containing additional amino acid sequences encoded by a viral structural gene. Is this enough to explain the disparate pathogenesis? (3) Payne *et al.*¹ and Cooper and Neiman (quoted in Hayward *et al.*⁴) also find evidence for ALV insertion near *c-myc* but in some cases the provirus integrates downstream of the cell gene and it is not clear how the viral promoters directly determine *c-myc* expression (Payne *et al.*¹ and P. Neiman, personal communication). The possibility of the virus acting as an insertional mutagen cannot be dismissed, particularly in view of the structural similarities between the integrated provirus and transposable genetic elements (discussed recently in *News and Views*⁵ and see Neiman *et al.*⁶). (4) Cooper and Neiman recently showed that DNA from ALV-induced lymphomas

can transform recipient mouse cells⁷. This transforming DNA contains neither viral structural genes nor viral LTR regions and they have now shown (quoted in Hayward *et al.*⁴) that the transforming DNA also lacks *c-myc* sequences.

Some resolution of these discrepancies can be attempted by postulating that *c-myc* functioning is needed for initiation but not for maintenance of lymphoid leukaemia. Provirus integration may stimulate *c-myc* expression (or perhaps maintain expression if the gene is already active) but this may serve simply to render the cells susceptible to further oncogenic change. A similar mechanism has been proposed for Burkitt's lymphoma in man, in which Epstein-Barr virus is thought to stimulate proliferation of B cells that are targets for additional neoplastic events, possibly involving chromosomal alteration. This view is supported by the natural history of avian leukaemia: although lymphomas are not seen until four to six months after infection, ALV-induced nodules can be seen in the lymphoid organ, the bursa of Fabricius, within six weeks. The status of ALV and *c-myc* expression in these nodules has yet to be fully investigated. The nodules seem to be clonal, but only a subset of the cells they contain may progress to become metastatic lymphoma cells⁸. It is mainly these lymphoma cells that were studied by Payne, Neel, Hayward and co-workers^{1,2,4}, and the failure to detect either virus or *c-myc* expression in some of the cells, and in all transforming DNA⁷, may reflect tumour progression unrelated to the events initiated by ALV infection.

Whatever the role of *c-myc* in neoplasia, the finding that ALV infection is regularly accompanied by the expression of the cell homologue of a viral oncogene forges a link between mechanisms of oncogenesis by transforming and non-transforming retroviruses. The concept that cell genes can be neoplastic when controlled by a virus promoter has been seminal in recent investigations on the genetic changes that occur after chemical carcinogenesis (discussed in a recent *News and Views*⁹). The possibility that non-transforming retroviruses activate similar genes adds coherence to oncogene hypotheses of tumour formation and provides another refined system in which to investigate such ideas. ALV oncogenesis will clearly be studied in yet more detail and parallel investigations are doubtless being pursued with other retroviruses that do not carry their own oncogenes. □

1. Payne, G. S. *et al.* *Cell* **23**, 311 (1981).
2. Neel, B. G. *et al.* *Cell* **23**, 323 (1981).
3. Quintrell, N. *et al.* *J. molec. Biol.* **142**, 363 (1980).
4. Hayward, W. S., Neel, B. G. & Astrin, S. M. *Nature* **290**, 475 (1981).
5. Flavell, A. *Nature* **289**, 10 (1981).
6. Neiman, P. E., Beemon, K. & Luce, J. A. *Proc. natn. Acad. Sci. U.S.A.* **78** (in the press).
7. Cooper, G. M. & Neiman, P. E. *Nature* **287**, 656 (1980).
8. Neiman, P. E., Payne, G. S. & Weiss, R. A. *J. Virol.* **34**, 178 (1980).
9. Rigby, P. W. J. *Nature* **290**, 186 (1981).

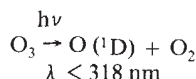
Measuring atmospheric gases and aerosols

from Stuart A. Penkett

APART from the formation of ions, there are two major photochemical processes which have been identified in the atmosphere. One results in the formation of a layer of ozone in the lower stratosphere and the other gives rise to a general oxidation process in the troposphere. The theory accounting for this second process, although put forward more than ten years ago¹, is not yet fully confirmed. It predicts that many chemical species such as hydrocarbons, halocarbons, sulphur and nitrogen compounds, are oxidized by hydroxyl radicals in the troposphere and converted to more soluble species which can be removed from the atmosphere by rain. This prevents many compounds, which are emitted both naturally and as pollutants, from reaching the stratosphere and interfering with the equilibrium of the ozone layer. The experiment entitled GAMETAG (Global Atmospheric Measurements on Tropospheric Aerosols and Gases) was an attempt by a group of American scientists to provide some evidence on the workings of this process. A series of twelve papers describing the work which took place in 1977 and 1978 has appeared in a recent issue of the *Journal of Geophysical Research*².

By the standards of atmospheric chemistry the experiment was a big one, based around the facilities provided by the National Center for Atmospheric Research at Boulder, Colorado (NCAR). It involved about forty scientists under the general guidance of Douglas Davis, C.S. Kiang, Paul Crutzen and Robert Duce and utilised a good proportion of the NCAR facilities for two years, collecting data over a flight track extending from Northern Alaska, down the North American continent and across the Pacific to New Zealand.

The most ambitious part of the experiment involved attempts to measure the hydroxyl radical concentrations both in the boundary layer and in the free troposphere above it. Hydroxyl radicals are formed by photolysis of ozone with light at wavelengths below 318 nm, which is about 30 nm above the minimum wavelength of solar radiation that penetrates the troposphere.



The excited oxygen atom which is formed reacts with the large amounts of water vapour present to give rise to hydroxyl radicals.

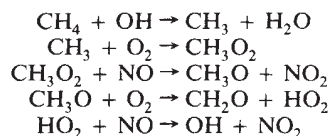
Stuart A. Penkett is in the Environmental and Medical Sciences Division, AERE, Harwell, UK.



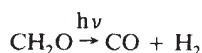
The concentration in clean air has been estimated to be about 5×10^5 molecules per cm^3 and various measurements in polluted air suggest a concentration range of 5×10^6 to 5×10^7 molecules per cm^3 (see refs 3,4). The best ground-based measurements have used a long pathlength UV absorption instrument which is not yet capable of measuring levels below 1×10^6 molecules per cm^3 . The fluorescence instrument used on GAMETAG was unlikely to improve on this, so although disappointing, it is perhaps not surprising that no data on the OH radical concentrations are presented.

Measurements were made on the global variation of the ozone and water vapour concentrations and of the photoproduction rate of singlet oxygen atoms; these enable the radical concentrations to be calculated by two-dimensional models of atmospheric chemistry which are currently being developed. In a world where measurements are not possible, calculations will have to suffice.

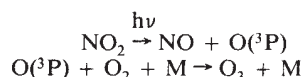
The radicals react with atmospheric trace gases such as methane in a complex chain reaction:



The formaldehyde is photolysed to give carbon monoxide and hydrogen



and NO_2 is photolysed to give nitric oxide and oxygen atoms.



The ozone which is formed in this way, and also ozone formed from CO oxidation, represents an alternative source to transport from the stratosphere⁵. The experiment showed that distinct ozone layers do exist in the troposphere, but in those observed over the ocean an inverse correlation between the ozone and water vapour concentrations indicated a stratospheric source. There was also no positive correlation between the ozone and carbon monoxide concentrations. The reason is almost certainly lack of sufficient nitric oxide to sustain the chain reaction. Recent measurements⁶ over the Pacific Ocean confirm this and the conclusion

must be that the extent of the atmospheric oxidation process over the ocean is very limited and far from producing ozone, photochemical reactions act as a sink.

Measurements of sulphur and halogen gases and many aerosol components both in and above the atmospheric boundary layer were also made. The sulphur measurements were particularly interesting since they suggest that sulphur dioxide is formed in the free troposphere above the boundary layer. There is also evidence that much of the sulphate aerosol observed was formed by a gas to particle conversion process occurring within the atmosphere. Atmospheric sulphur chemistry has been a subject of discussion for three decades and more. Much of this has focused on the natural sources of sulphur, whether they are present as hydrogen sulphide or dimethyl sulphide. Recently two more gases, carbonyl sulphide and carbon disulphide, have been shown to be present in the atmosphere⁷ and it has been suggested that these are the main precursors of sulphur dioxide and the sulphate aerosol found in the remote troposphere⁸. The GAMETAG experiment has certainly proved beyond doubt that carbonyl sulphide is the most abundant sulphur-containing trace gas in the atmosphere. However, it is not clear yet how quickly it is oxidized^{9,10}.

GAMETAG showed very little evidence for the presence of long-range transport of pollutants from the continents over the ocean. This was true of gases such as carbon monoxide and sulphur dioxide and for aerosol species such as sulphate and nitrate. It is also clear that the ocean is the major source of methyl chloride, which is the only naturally produced halogen-containing gas to enter the stratosphere in large quantities.

In conclusion, GAMETAG provided some interesting data on the concentration distribution of stable trace gases and aerosols and the influence of pollutants on the composition of the remote troposphere. However, it failed to provide positive evidence for the existence of a large amount of chemical activity in the troposphere. Most probably this is associated with low levels of nitric oxide over the oceans, which suppresses the level of hydroxyl radicals.

1. Levy, H. *Science* **173**, 141 (1971).
2. *J. geophys. Res.* **85**, No. C12, 7285 (1980).
3. Perner, D. *et al. Geophys. Res. Lett.* **3** 466 (1976).
4. Wang, C.C. *et al. Science* **189**, 79 (1975).
5. Fishman, J. and Crutzen, P.J. *Nature* **274**, 855 (1978).
6. McFarland, M. *et al. Geophys. Res. Lett.* **6**, 605 (1979).
7. Sandilands, F.J. and Penkett, S.A. *Atmos. Environ.* **11**, 197 (1977).
8. Logan, J.A. *et al. Nature* **281**, 185 (1979).
9. Kurylo, M.J. *Chem. Phys. Lett.* **58**, 238 (1978).
10. Cox, R.A. and Sheppard, D. *Nature* **284**, 330 (1980).

The continents are likely to be the major source of nitrogen oxides and these are the places to look for evidence of atmospheric reactions. There is certainly a case for an experiment of this type to continue if only to confirm that many compounds released from modern industry are destroyed before they can reach the ozone layer in the

stratosphere. It should concentrate on unpolluted continents in tropical latitudes and it should include measurements of transient products such as aldehydes and nitro compounds. These are easier to measure than radicals but provide the sort of information which is necessary to test theories of atmospheric oxidation. □

closing of pre-existing cracks.

How then do induced cracks attain the characteristics of pre-existing cracks? A key element in this seems to be the extent to which cracks are modified by the circulation of water at high temperatures. Under these conditions grains become rounded and crack walls etched and pitted (Potter, J.M. *Rep. LA-7224-T*, Los Alamos Sci. Lab., 1978) and begin to resemble pre-existing cracks.

The fact that almost complete closure of thermal and stress-induced cracks can occur lends support to D.J. Holcomb and J.L. Stevens' assertion that reversible Griffith cracks can explain all of the observed attributes of rock dilatancy (*J. geophys. Res.* **85**; 7101, 1980). They cast doubt on the widely used sliding crack model of dilatancy, in which dilatancy arises from the formation of tensile cracks at the ends of a large number of sliding shear cracks, partly because a shear crack model cannot explain the stress memory rocks exhibit and especially because shear cracks are only rarely observed in some dilatant rocks (Tapponier, D. and Brace, W.F. *Int. J. Rock Mech. Mining Sci. Geomech. Abstr.* **13**; 1976). On the other hand, long, narrow cracks preferentially oriented parallel to the maximum compressive stress are very common in some rocks and these Griffith cracks can apparently explain all of the experimental results on rock dilatancy, provided that the cracks can open and close reversibly under stress.

A problem arises, however, because most studies of stressed cracks in rocks have been made on low-porosity, well-cemented crystalline rocks. Less well-cemented, high-porosity sedimentary rocks with more point contacts and fewer interlocking grains have quite different initial preferred orientations and distribution of microcracks. Recent experimental studies have shown that the differences in peak strength, the localization of deformation to give a main fault, and some water-weakening effects can be qualitatively explained in terms of the initial microstructure (Hadzadeh, J., Thesis, University of London, 1980).

Two limitations of the MIT study are that we are not sure their observations are typical of the bulk of their specimens. They examined sections that were within 1 mm of the surface, and the role of stress corrosion crack growth was not investigated in the high-vacuum environment of the SEM. *In-situ* tensile tests, however, show that stress corrosion can be a significant factor (Swan, G. *Res. Rep.*, Tulea 1980: 01, University of Lulea).

It would now be valuable to extend *in-situ* observations to higher stresses and temperatures and more complex polyaxial stress states. At present this is not possible and one is forced to examine at zero stress and ambient conditions specimens that have been retrieved from more exotic conditions. □

Cracks in rocks under stress

from Barry Atkinson

MICROCRACKS influence many of the physical properties of rocks. Strength, transport, elastic and inelastic characteristics are all highly dependent upon the size, shape, number and distribution of microcracks. These physical properties change dramatically as crack characteristics are modified under stress.

In earthquake prediction, the exploitation of oil, gas and water bearing formations, nuclear waste disposal, and heat extraction and storage, the prediction of *in-situ* physical properties from models of rock microcracks is of great importance but has, so far, been rather unsuccessful. The basic problem is that, despite many laboratory studies, we have only a rudimentary idea of the actual modifications that cracks undergo in stressed rocks. M.L. Batzle, G. Simmons and R.W. Siegfried of MIT have recently attempted to overcome this problem by making direct observations of crack closure in granite and diabase under uniaxial compressive stress (*J. geophys. Res.* **85**; 7072, 1980).

The MIT group placed a small hydraulic press inside a conventional scanning electron microscope (SEM) so that they could subject cylinders about one cm in diameter and two cm long to uniaxial stresses of up to 30 MPa. A portion of the lateral side of the core was ground flat and bombarded with ionized argon to remove the damaged surface before examination in the SEM. Stress-strain measurements were also made on similar cores in the press but outside of the SEM. Although SEM studies have been made before on specimens subjected to tensile *in situ* stresses, this is the first time that a compressive stress has been so studied.

Westerley granite and Maryland diabase were examined in the 'natural' state and after slowly heat cycling to 500°C and 700°C, respectively, to increase crack porosity. Westerley granite is a typical crystalline rock with initial crack porosity of about 0.2 per cent and a total porosity of about 1 per cent. Maryland diabase is unusual in that it has no measureable crack porosity and a low total porosity of 0.1 per cent. Westerley granite consists of roughly equidimensional feldspar, quartz and

biotite mica grains averaging 800 micrometres in diameter. Maryland diabase consists mainly of equidimensional grains of feldspar and pyroxene averaging 200 micrometres in diameter.

In general, the response of cracks to stress confirmed those predicted from microscopic examinations made at zero stress (Kranz, R.L. *Int. J. Rock Mech. Mining Sci. Geomech. Abstr.* **16**; 37, 1979) and from deformation measurements made under applied stress (Brace, W.F. *et al. J. geophys. Res.* **71**; 3939, 1966). Numerous factors influence the behaviour of cracks, including orientation, shape, source, and the proximity of other fractures.

The main conclusion of the recent work, however, is that pre-existing cracks in these rocks behave quite differently to those induced by heat cycling. Pre-existing cracks have walls that are irregular, etched and pitted, and poorly matched. Under stress, closure is incomplete with many portions remaining open and interconnected. The shape, or aspect ratio, continually changes as stress increases. Thermally induced cracks are also irregular but the walls are more perfectly matched and tend to close more completely, unless propped open by very small grains or debris.

Although pre-existing fractures can be re-opened, widened, extended, or closed, depending on a host of interrelated factors, in general those cracks oriented close to the direction of maximum stress tend to open, whereas those oriented nearly perpendicular to the maximum stress tend to close.

New cracks created in the laboratory by stress or temperature usually have the irregular but well matched characteristics of thermal fractures. This is important because the precise nature of the crack response to stress determines the trends in dependent physical properties. New fractures will close easily and hence lead to a large change in physical properties with stress as compared with the more inefficient

Barry Atkinson is NERC Special Research Fellow at the Geology Department, Imperial College, London.

SPRING BOOKS SUPPLEMENT

Reviewers without pride or prejudice

What are book reviews for? The question is not entirely trivial. The pages that follow are filled with the comments of a great many serious people who have taken the trouble to read some book or books, to reflect on the contents thereof and then to tell the rest of the world what they think of them. Why, it will be asked, should they bother? Why should an army of people who are (or who should be) disinterested take this trouble for the sake of others — the people who lie somewhere between those who would certainly buy the book concerned, perhaps because it represents a special interest of theirs, and those who would not willingly be seen in public reading the document concerned, and who would not wish to do so in private either? Pecuniary reward is not the explanation. The simple answer is that book reviewers pursue one of the most honourable of intellectual crafts. The ideal reviewer is a self-effacing person — the kind of person who is sufficiently knowledgeable to have written the book he is charged with off his own bat, who might well indeed have done so if he were not so busy, but who has nevertheless been willing to quarry time from a busy timetable to inform and instruct potential readers of a new apparition in the literature. The scientific community, any intellectual community, has reason to be grateful to reviewers.

But why are books so especially deserving of attention? What is it about a pair of hard covers that gives a book precedence over, say, an article in a scientific or other journal? Rarity is only part of the explanation. Some journals, *Nature* among them, do indeed try to pick out and draw to the attention of their readers particularly interesting articles that appear in the ordinary course of communication. The volume of the scientific literature is of course too great for this task to be attempted (or at least accomplished) systematically. There is, however, another sense in which a book commands (and deserves) to be dealt with differently from the report of some piece of research, whatever the excitement engendered by the results. For a book, or at least a good book, is necessarily a somewhat personal declaration. So much can be told from a comparison of the style of run-of-the-mill scientific articles and books by the same authors. The authors of articles run to the passive, and to opening sentences such as "There is now much evidence to suggest that . . .". The authors of books write in a different frame of mind. They are, from time to time, prepared to utter sentences beginning "I believe that. . .". More than this, books differ from articles in the journals in their objectives. Both kinds of contributions to the literature have a beginning and a middle, but only of books (or good books) is it required that there should also be a conclusion of some weight and finality. The authors of articles in the journals are allowed to finish with a statement to the effect that they are entirely at a loss to understand the observations they report and to declare that they will do their best, in competition with those who are stimulated by reading their contribution to the literature, to find out. Would-be book authors, by contrast, must (or should) know before they put pen to paper what case they seek to make. (Publishers who make books by stringing together a miscellaneous collection of articles, or collections of papers given at some symposium, must by the same token appreciate that these are comparatively dull, not much like books.)

There is another crucial respect in which a book differs from some other contribution to the literature. Although some books may only be ten times longer than, say, some journal articles, the

difference of scale is a quantitative difference big enough to imply a qualitative difference. As the world knows, book authors are repeatedly perplexed to know whether the point they are about to make in chapter umpteen has already been made in chapter one. Put simply, those who set out to write books, and actually finish them, command a certain respect, even from those who may disagree with the conclusions.

Good reviewers appreciate these truths, and are (or should be) compassionate people. If somebody writes a routine account of, say, the inorganic chemistry of selenium, and if the arrangement of the chapters is not precisely what the reviewer would have chosen, that does not amount to an excuse for a fierce attack on the author and his good sense and a warning that students everywhere should avoid the book like the plague. In such fields it is well known that the unkindest course a reviewer can follow is to say nothing at all: then potential purchasers will have to learn of the book by chance, while the book's publishers will be at their wits' end, seeking other ways of putting news of the book's appearance into circulation. In the last resort, after all, the reviewer's task is to tell potential readers what a book seeks to accomplish and to say whether the author has done what he set out to do. Part of the book reviewer's licence is that he may, in passing, add a few reflections of his own on the good sense of the author's goal. So most book reviews should be essays in sympathetic understanding, marked sometimes with sorrow, rarely with anger. And so they usually are.

There are, however, exceptions to this benign rule. For the class of well-conceived books, each with its own sharp conclusion, is bound to include many whose conclusions are at once mistaken, even mischievous, but also likely to be persuasive. This is why the craft of book reviewing frequently runs to polemic. This is why H. J. Eysenck's *The Causes and Effects of Smoking* (Temple Smith, 1980), or Samuel Epstein's *The Politics of Cancer* (see *Nature* **284**, 297; 1980) persuaded reviewers to soak their typewriter ribbons in vitriol. The arguments, right or wrong (and probably wrong, or at least incomplete), are such a challenge to accepted doctrine and at the same time so self-consciously challenging that it is helpful to readers to know the strength of opinion on the other side. Polemic is therefore part (but ideally a small part) of book review columns. Professor Dorothy Hodgkin's review of Maurice Goldsmith's book on J. D. Bernal (*Sage*, Hutchinson, 1980; see *Nature* **289**, 99; 1981) is more puzzling. Goldsmith set out to write a biography of one of his heroes, and was savaged for his pains by another hero-worshipper, his reviewer. No doubt the book is, as Professor Hodgkin says, "confused and inaccurate". It is also, however, a good book in the sense of being a good read and also (for a hero-worshipper) honest in that it deals with the warts as well as the achievements — sexual proclivities, ambivalence about Lysenko and all that. Maybe Goldsmith's book is "not the book that is needed about Bernal", but it is the only book we have. On this occasion, polemic seems to have been misplaced. In general, the truth about books is that they are milestones for somebody — certainly the author and the publisher — and that there are some occasions on which even bad books are better than no books. Moreover, books differ from other contributions to the literature of scholarship in the diversity of their readership. Their influence on students is important. And books — or good books — can survive from one decade to another.

Advice and dissent in White House science

Harvey Brooks

THE growing literature of science policy generally includes two kinds of books: first-hand accounts of their own experience by principal actors in the processes of high-level scientific decision-making, such as those by James Killian, George B. Kistiakowsky and Herbert York; or analytical discussions by academic scholars based on study of official documents or interviews with the principal actors. Edward Burger attempts to build a bridge between these two genres. A member of the staff of the Office of Science and Technology (OST) from the time of the first Nixon administration to the abolition of the President's Science Advisory Committee (PSAC) and OST in 1973, Dr Burger has built his book firmly on his own first-hand experience as an insider, but at the same time has attempted to bring a historical and analytical perspective to this experience and to put forth some important generalizations.

To the average observer of science policy, the Nixon administration tends to be viewed as the low point in the history of science in the White House. The rather sudden decision to abolish the presidential science advisory apparatus in January 1973 overshadows the many constructive initiatives that arose during this period. As Burger reminds us, this was a time when the effort to "civilianize" OST and PSAC — to bring scientific considerations and the potential of science fully to bear on some of the nation's major sociopolitical issues such as health care, environmental regulation, population policy and urban policy — reached its climax.

His account is notable for its political sophistication and sensitivity to the multiplicity of influences bearing on even the apparently scientific decisions. It is also refreshingly free of self-justification or overstatement of the role of science. The chapters on national health policy, on environment and health, and regulation to protect health and on population and family planning contain a wealth of new material relating to major Nixon administration initiatives in social policy which gradually became diluted by pluralistic political pressures. These chapters are well worth reading even by those who care little about science policy or the role of science in the White House.

The subtitle of the book is unfortunately somewhat ambiguous and fails to do justice to all the themes which Burger develops. Its meaning is never fully explained and conveys implications that are not really supported by Burger's detailed conclusions. It is thus not clear whether he means to imply that having a Science Advisor in the White House is a

Science at the White House: A Political Liability. By Edward J. Burger Jr. Pp.208. ISBN 0-8018-2433-8. (Johns Hopkins University Press: 1980.) \$14.95, £9.

political liability to the President, which tends to tie his hands in negotiating with the many contending constituencies that he has to satisfy, or whether having a prominent presence in the White House is a political liability for the scientific community which compromises its long-term credibility before the public. Perhaps both statements are true, but I do not believe that the implication that science should be taken out of the White House is either intended by Burger or justified by the case histories which he analyses so perceptively.

In fact one could argue that science in the White House enjoyed some of its finest hours during the period covered by this account, if retaining the intellectual honesty of science in the policy process is used as a criterion. The report on *Chemicals and Health*, published by the National Science Foundation after the demise of PSAC, is still one of the most thorough and balanced discussions of this controversial topic. Issued in a climate of overwhelming environmental activism, its sober and balanced approach was little appreciated or heeded until much later. Had it been heeded, some of the current over-reaction against the earlier excesses of the environmental movement might have been avoided. The Toxic Substances Control Act (1976) derived from a concept that originated from the staff of the Science Advisor. The OST staff, and Burger in particular, played a central role in shaping the recommendations of the Health Policy Review Group (1971) and the subsequent Presidential Message on Health. The analysis of the Review Group stands up very well today, though it had little political effect at the time. The OST also played an important role in following up the report of the Population Commission (1972), although here again the analysis was ahead of its time in political terms. The Science Advisor played a reluctant but sensible role in devising the Quality of Life Review Process for considering the social and economic impacts of regulatory decisions, anticipating the regulatory review process established in the Carter administration. The Science Advisor chaired a special committee created by the President to study the rather grandiose and politically motivated proposal for a New Technology Opportunity Program in 1971.

The recitation of these examples, however, poses the dilemma suggested in the title of the book. In referring to these

activities favourably I am making an intellectual — not a political — judgement, because in most instances the input from the OST, as Burger suggests, must be rated a political failure. The intellectual conclusions were mostly out of tune with the political climate and the President's political goals — the classic problem of "speaking truth to power". The report on *Chemicals and Health* took too long to complete and appeared too late to have any significant impact on policy, and its balanced approach was out of tune with the public's mood. By the time the hard hitting and clever recommendations of the Health Policy Review Group had been distilled into a presidential message, much of the bite had disappeared, and health as a salient political issue had faded into the background. Much the same fate befell the report of the Population Commission. The task force to follow up the report completed its work at a time when political concern over US population growth had faded with the dramatic decline in fertility in the early 1970s. Although the task force called for a presidential statement to endorse the Commission's principal recommendation for a national population stabilization policy, the chairman privately sabotaged this recommendation. The technology initiatives programme, which was launched with great political fanfare, quietly foundered when Dr Edward David's standing committee discovered there was much less there than met the eye, and was unable to identify any dramatic government initiatives that stood up to scientific scrutiny.

We thus end with a picture in which strong recommendations based on scientific input failed politically because they ran counter to the prevailing climate or antagonized strong political interests, on the one hand, or in which careful analysis debunked political initiatives on which the President had set great store, on the other. Are the Science Advisor and his staff to be praised for "speaking the truth to power", for calling the shots as they saw them and for thoughtful and forward-looking policy analysis. Or are they to be criticized for insufficient awareness of political realities and of the sensitivities of important political constituencies to which the President must respond? Is the performance of science in the Nixon White House to be viewed as a series of honourable failures derived from commendable devotion to intellectual honesty and objectivity? Or is it to be condemned as the mark of politically naive amateurs incapable of appreciating and responding to the legitimate political needs of the chief executive? Burger poses these



Scientists at the White House, 1957–1981. Left to right, top to bottom: James Killian Jr (served under Truman, 1957–1959); George Kistiakowsky (Eisenhower, 1959–1961); Jerome Wiesner (Kennedy, 1961–1964); Donald Hornig (Johnson, 1964–1969); Lee Du Bridge (Nixon, 1969–1970); Edward David Jr (Nixon, 1970 to 1973 when PSAC and OST were abolished); H. Guyford Stever (Ford, 1976–1977); Frank Press (Carter, 1977–1981). President Reagan has not yet appointed a successor to Press.

questions cogently in each of the examples which he analyses, but he does not really give us the answer. Perhaps there is no single answer and the greatest contribution of the book is revealing the dilemma in realistic detail.

An important theme of this book, set forth in Chapter 2, is the fundamental conflict between rational policy planning and the push and pull of pluralistic political interests. The author presents the executive branch and the scientists as the proponents of such rationality against the pluralism embodied in the legislature. Yet there is a paradox here, which is not fully resolved in the book. It is true that in the efforts to introduce Program Policy Budgeting (PPB), to develop a comprehensive health policy or to formulate a rational and balanced national policy for environmental regulation, Burger's model seems to apply. But he fails to note that in policy for science, as contrasted with science for policy, the roles of the scientists and politicians appear to be reversed. It is Congress that has continually pushed for a comprehensive national policy for science, most notably in the National Science and Technology Policy, Organization and Priorities Act of 1976. Furthermore, this is

not a new phenomenon. In the nineteenth century, Congress was the repeated source of proposals for a cabinet Department of Science. In every case it has been the scientific community and the executive branch that have dragged their feet in regard to the comprehensive planning of science and technology. More recently, OSTP under the Carter administration has attempted to divest itself of most of the planning and systematic reporting functions mandated by the 1976 Act. Whereas Congress has viewed OSTP and the Science Advisor as a kind of general staff for science and technology in the United States, the Science Advisor and most of his constituency in the technical community have viewed his role more modestly and have strongly favoured a pluralistic, decentralized approach to the support of science itself.

It is interesting to speculate on the causes of this paradox of science as the advocate of planning in the spheres of public policy and pluralism in the field of science, while the Congress and the more politically minded Executive take the exactly opposite view. The answer may lie in the internal politics of the scientific community itself. This community is not a monolithic group

but rather congeries of frequently conflicting interests and views. Planning for science, especially by scientists, entails a high political cost within the scientific community, which would rather take its chances on the free play of political advocacy between the different segments of the scientific community and the political process. It would have been desirable if Dr Burger had explained this particular paradox a little further.

Altogether, though, this is a valuable book which should be of interest to a wide public. It provides some fascinating insights into some of the major social policy debates of the early 1970s; it is a contribution to the analysis of the role of White House science advice from the perspective of a staff officer carrying out policy support rather than making policy; and it provides excellent case material for scholars and students in the field of science policy. On top of that it is well written, literate and sophisticated. □

Harvey Brooks is Benjamin Pierce Professor of Technology and Public Policy at Harvard University. He was a member of the President's Science Advisory Committee from 1959 to 1964.

Education and the compleat scientist

J. Z. Young

Teaching and Learning about Science and Society. By John Ziman. Pp.190. ISBN 0-5212-3221-X. (Cambridge University Press: 1980.) £9.50, \$22.50.

MOST scientists are interested in scientific education, if only for the memory of their own debt to it. I imagine that a great many became scientists because they were lucky enough to have an inspiring teacher. My own at Marlborough College was A. G. Lowndes, a remarkable man who trained several Fellows of the Royal Society and a Nobel Prize Winner (Peter Medawar). He inspired us by his own somewhat amateur researches, he showed us endless living things and, by making us work even in our spare time, taught us that all science involves much hard work.

These are perhaps the main items that must be included in any scientific education: it must catch enthusiasm, it must cover a huge mass of facts and it must find out those who can enjoy learning a lot of detail. This is how scientists are made, but in the process they seldom learn much about the position of science in society, still less about its history or philosophy.

It is these defects that are the particular concern of Professor Ziman as indicated by the title of his book. I cannot remember ever hearing them discussed by my teachers either at school or university. Probably many scientists still take up their subject because they are interested and like it, though there is now more concern about its social implications. They are all too liable to fall into what Ziman calls "scientism and technocracy", the uninformed assumption that everything that is done in the name of science is good. How can this restriction of the development of young scientists be avoided? Ziman has tried to encapsulate the problem under the acronym STS — Science, Technology and Society. These themes permeate the political, economic and cultural issues of our times yet "there is no *rationale* of STS as an educational subject", although various courses of study of it have been tried. In trying to provide this rationale Ziman manages to stimulate thoughts about a great many fundamental questions, even beyond those dealing with the educational problem itself.

For instance, one feels that the discussion obviously needs a definition of science, and many times Ziman seems on the point of providing one, but wisely never comes down to calling it a definition. Sometimes it is a question of method: "Science derives its practical power and authority from the rigours of its arguments and the hardness of its facts". Moreover, he repeatedly stresses that the production of scientific knowledge is a social process. Then only a page later we are in a different world of discourse: "There is no single

'scientific' map of reality — or if there were, it would be much too complicated and unwieldy to be grasped or used by anyone". Is this true? I suspect that Ziman doubts it himself. He pursues the question into a discussion of the hierarchy of scientific enquiry and the questions of emergence and reductionism,

the notion that the properties of complex systems such as organisms or molecules can be 'reduced' to the laws satisfied by simpler systems such as cells or atoms — is not only a very dubious philosophy, it is a dangerous folly in science education, where the map appropriate to each level must be taught wholeheartedly according to its own lights

Yet probably most scientists have a hankering for that "folly". And is it indeed so foolish? It is true that chemistry and biology have their own laws but no one can go all the way in either of those sciences unless they know something about physics. However much "autonomy" there may be for each part of science there is still, at least for many of us, a scientific map or model of the world to which we try to refer all events and all knowledge, including the knowledge of ourselves. It seems that the human brain is so constructed that it tries to build all its information and schemes of action around a unified model. Of course this grows gradually in each one of us, centred at first around a parental scheme. The scientific model, if properly learned, can provide the adult substitute, which many people feel they lack in the absence of religion. That is of course not to say that science should be dogmatic or be treated as religion, but that it can provide what Ziman calls "the possibility of discovering order in nature" and to do this is a requirement for human life. We all need some system of order. Incidentally this is a characteristic specific to human beings and not present in our nearest relatives. Ziman actually recognizes this later in his book

when he says that "scientific world pictures" allow "deductions that help us along the way through life".

This is only one of the ends that he believes could be achieved by the proper attention to STS education. He holds that they can only be reached by specifically designed courses for all levels of scientific education. In the later part of his book he discusses how these should be planned for different stages. He is against "General Science that is too sloppy and technical science that is too arbitrary", and he advocates a "valid science". This is easy enough to say until you come to the job of deciding what to put in and what to leave out. Ziman discusses the many ways in which it could be done. For instance, it can be through the social relevance and applications of science, through its possibilities as a vocation for the individual, or its history, or its philosophy or its value in solving world problems. But he does not try to specify the content of particular courses. His aim is to show that the teaching of science at present is grossly defective because it does not properly develop the understanding of the individual either of himself or society, or indeed of the nature of the science he is learning. His plea is that teachers of science should take definite steps to fill this gap by designing what he calls STS courses. He knows that there will be much opposition and anyone who has tried to introduce such courses will have experienced this. But there are many people ready to agree with his plea that a better understanding of the relations of science and society is needed for the proper training of research scientists as well as doctors and other technologists, not to mention civil servants, politicians and the general public. A major value of this interesting book is that it makes you realize how difficult it is to reach agreement either about the nature of science or its proper place in society. [1]

J. Z. Young is Emeritus Professor of Anatomy at the University of London.

IQ gladiators in separate combat

Stuart Sutherland

Intelligence: The Battle for the Mind. H.J. Eysenck versus Leon Kamin. Pp.192. ISBN hbk 0-333-31279-1/0-471-08884-6; ISBN pbk 0-330-36399-4. (Macmillan Press, London/Wiley, New York/Pan: 1981.) Hbk £12, \$12.95; pbk £2.95.

NATIVISTS have battled with empiricists for many years, but never with such fury as over the inheritability of intelligence. It might be thought that the issue could be resolved by bringing together representatives of the opposing sides and letting them argue out, point by point, the

evidence for the respective roles of inheritance and of environment. *Intelligence: The Battle for the Mind* could have provided just such an opportunity, but unfortunately it is constructed in such a way that the combatants only skirmish and never meet in a decisive battle.

The book contains four sections. In the first, Professor Eysenck sets out his case for supposing that in Western society inheritance accounts for about 80% of the variation in intelligence. Although he does not enumerate them, he deploys 20 different lines of evidence from which he claims this inference can be drawn. In the second

section, Professor Kamin produces his criticisms of some of the evidence for the importance of heredity, but this section is not a reply to Eysenck's initial foray since it was evidently written without Kamin having read it. In consequence, many of Kamin's arguments are beside the point: for example, although Eysenck excludes all reference to Burt's data, now universally acknowledged — thanks to Kamin's brilliant detective work — to have been faked, Kamin devotes a chapter to attacking Burt. In the final two sections, which are much too brief, Eysenck replies to Kamin's set-piece and Kamin to Eysenck's, and there the matter rests.

No agreement is reached on which aspects of the evidence are reliable and which are not, and both authors are guilty of failing to meet the arguments adduced by the other. Thus, in his opening section Eysenck ignores many of the criticisms previously levelled by Kamin. For example, one of the strongest arguments in favour of the role of inheritance is that when monozygotic twins are brought in by foster parents, there is a higher correlation between the IQ of the natural parents and the child than between that of the foster parents and the child. Kamin has put forward two ingenious suggestions to explain this correlation in environmental terms. First, many of the foster children in question spent a year or more of their lives with their natural parents before moving to foster parents. Second, the IQ and the socio-economic status of foster parents is in general high and has much less variance than the IQ of the population at large: this reduction in environmental variance would reduce the effects of environment. One would have liked to have seen Professor Eysenck's reply to these arguments, but it is not to be found in the book. On the other hand, Eysenck points out that there is a higher correlation between the IQs of monozygotic twins reared together, but one looks in vain for Kamin's response to this argument.

Both authors give the impression of being determined to make a case and of selecting data that forward that case. It is a pity that Eysenck did not concentrate on the most solid evidence for inheritability and ignore dubious or unrepeatable findings. For example, he argues that there is a difference in the brain waves (EEGs) of the intelligent and unintelligent and that this difference suggests intelligence is inherited. Not only is the argument fallacious, since there is no reason why brain waves should not be in part determined by environmental factors, but recent attempts to repeat the finding Eysenck uses have failed.

Eysenck points out that Kamin does not have a theory of intelligence. With great ingenuity, Kamin has taken a number of pieces of evidence thought to support the hereditary case and has shown that there are other *possible* explanations: he has not proved that these other explanations are

correct. Moreover, since Kamin treats each piece of evidence in isolation, he is able to use individual arguments that are inconsistent with one another. Thus, as Eysenck notes, Kamin explains the large differences in IQ between dizygotic twins reared *together* by assuming a large difference in the way they are treated within the family. Elsewhere, however, he explains the similarity in the IQs of monozygotic twins brought up *apart* by arguing that each pair is placed in families of a similar socio-economic background. Taken together, Kamin's two arguments imply the absurd conclusion that there is less environmental difference *between* families than *within* a family.

There are a few things on which both authors agree. Moving children from deprived homes to particularly good homes can bring about a shift of up to 20 points in IQ. Although Kamin regards this as compelling evidence for the overriding importance of environment, Eysenck is able to show that it is consistent with his own theory of intelligence. He does not maintain that a large shift in environment cannot change IQ, merely that in conditions as they exist today IQ is more determined by heredity than environment.

Kamin does not help his case by his personal virulence towards Eysenck. The fact that Eysenck has twice made a mistake about the sex of an author whom he cites has surely nothing to do with the inheritance of IQ; and it is merely distracting to the reader for Kamin to stress past errors made by Eysenck, at least some of which (for example, his defence of Burt) Eysenck now acknowledges. Moreover, Kamin's allegation that the aim of the science of genetics is to make "the world comfortably safe for white males" is both vituperative and false — white males were a good deal safer before IQ tests were ever thought of. Eysenck rightly insists that differences in the average IQs of groups have no bearing on how the individual should be treated. It is impossible to predict the individual's IQ from skin colour, sex or social class and in at least some instances the use of objective tests for selection has favoured underprivileged groups. Burt may have been a scientific scoundrel, but his introduction of intelligence tests for secondary education in Britain between the wars doubled the proportion of poor children in secondary schools.

Like all scientific discoveries, the results of work on intelligence testing can be used for good or ill. As Eysenck notes, recent decisions by anti-racists in the United States are likely to be extremely damaging to minority groups. For example, it is folly to insist that equal proportions of blacks and whites should enter training programmes for the educationally subnormal, if, for whatever reason, a higher proportion of blacks than of whites need the help of such training. Eysenck quotes a remark of Dr Johnson's which sums up the position: when asked whether

men or women were more intelligent, he replied: "Which man? Which woman?"

Intelligence: The Battle for the Mind is a wasted opportunity. One feels that had Eysenck concentrated on the most compelling arguments for the role of inheritance, and had he attempted from the outset to meet Kamin's ingenious but *ad hoc* arguments where they could be met and conceded cases where they could not be met, the debate might have been settled. Eysenck might not have proven that IQ is 80% inherited, but he might have established once and for all that there is a strong inherited component.

Neither combatant comments on the scientific importance of intelligence testing. A much more destructive criticism can be made of the whole enterprise than any advanced by Kamin, namely, that it has told us nothing about how the human mind works nor has it given any answer to the really important applied problem — how can we make it work better? □

Stuart Sutherland is Director of the Centre for Research on Perception and Cognition at the University of Sussex.

Reality of science

Peter Newmark

An Imagined World. By June Goodfield. Pp.288. ISBN 0-06-011641-2. (Harper & Row: 1981.) \$12.95 US only.

IF A scientist wishes to discover what makes a certain type of cell tick, she may well spend five years studying its behaviour and composition. If a writer wishes to find out what makes that scientist not only tick but chime, by way of making scientific discoveries, she may decide to track the scientist's every thought throughout the five years. That neither venture is guaranteed to succeed is illustrated by this book. The story, however, is not without interest.

In 1975 June Goodfield met Anna Brito (a pseudonym) who was then on sabbatical in New York and, within hours, decided that she fitted the bill of an individual research worker through whom to follow the process of science. Goodfield found Anna Brito "articulate, amusing and somewhat different from the majority of scientists I had met before", and sensed that Brito was at a stage where doors were beginning to open. In one sense, at least, she was right. Within a year of returning from New York to Glasgow University, the doors of New York's Memorial Sloan-Kettering Cancer Center had opened to Anna Brito. From then on Goodfield had ready access to the scientist. So except for the first year and for a period in 1977, when progress is recounted through letters and tape recordings, the tale is told in narrative form.

Through the letters, recordings and narrative emerges a story that illuminates the working life of a scientist in a detailed and open way that has seldom, if ever, been done before. Unlike *The Double Helix*, *An Imagined World* contains little glamour, much toil, many failures, much anguish and no prizes. It is, in a word, realistic. Whether that makes for interesting reading is a matter of taste. Certainly for the working scientist there will be few surprises, but there may be reflections, both pleasant and not so pleasant, to be seen in the mirror of Anna Brito's progress. For the non-scientist who does not realize just how great a travesty of the making of a discovery is the usual account given in a scientific paper, the book is bound to come as an eye-opener.

It would be hard, however, to claim that Anna Brito's experiences were particularly typical. Her first, and arguably most important, discovery certainly came about in an unusual way. She found herself in a London laboratory, unable to speak much English, having neither skills nor ideas to offer and altogether a rather unwanted visitor. Her supervisor, somewhat in desperation, gave her a large collection of sections of mice to examine through a microscope. Against all the odds she noticed something both original and interesting. Her observation amounted to the discovery that T-lymphocytes and B-lymphocytes have predilections for separate zones of tissues.

In the years that followed, Anna Brito pursued her discovery of what she called "ecotaxis" and developed the idea that certain diseases were associated with wrongly located lymphocytes. This condition she called "ecotaxopathy", a term guaranteed to turn purple the faces of many immunologists. Clearly not a lady to be thrown off course by such reactions or by what, to many, would surely be an unnerving number of grant refusals, her research continued in the same vein throughout the period covered in the book.

Much of what is recounted rings true; for example, the long periods Anna Brito spends reading around her subject hoping to place her observations into a broader context. Other features of her laboratory life seem less convincingly told. Only twice, for example, is there any mention of her having any scientific competitors. This is either true and untypical, or untrue and misleading, given the extent to which competition drives so many scientists.

The chief virtue of the book is that it recounts in detail the thoughts and actions of a scientist brave enough to be open about her mistakes and failures and to bare many of her thoughts, despite the risk that some might look naive at the time and worse in retrospect. But the book fails to shed much light on the process of scientific discovery or to sustain interest for one simple reason; as it happens, nothing that I could classify as a substantial discovery emerges during the period covered in the

book.

Finally, potential readers might be attracted by two puzzles. The first, not too taxing, is the real identity of Anna Brito. The second, perplexing, is why, given the circumstances of the initial discovery that set her on the pilgrimage of research, Brito

can later say: "I can't stand people who just put everything together and try to see what happens — who don't think before doing an experiment . . . I could never do science like that?"

Peter Newmark is Deputy Editor of Nature.

Will the real Beaker Folk please stand up?

Colin Burgess

The Beaker Folk: Copper Age Archaeology in Western Europe. By Richard J. Harrison. Pp.176. ISBN 0-500-02098-1. (Thames & Hudson: 1981.) £12, \$19.95.

"BEAKER Folk", like the Wee Folk, exist in the mind of the beholder, and Dr Harrison is as uncertain in his attitude to "Beaker Folk" as most of us are about fairies. His last chapter, "Was there a 'Beaker Folk'?", was eminently sensible, and his conclusion that they "have no substance as a special population group", and that the Bell Beaker phenomenon began "as a fashion in fine pottery", would meet with widespread approval among students of the subject.



Gorge of the River Verdon in Provence, "an exceptionally interesting site".

Unfortunately, this attitude does not square with the contradictions in his preceding six chapters. The author claims in several places not to believe in "Beaker Folk", but clearly finds it difficult to erase them from his subconscious. His inner confusion will leave the reader feeling profoundly uncomfortable, because it is obtrusive throughout: to quote from the jacket, "the so-called 'Beaker Folk' were not a separate, well-defined community, but constituted more of an influence or movement", yet in the next sentence "They appear to have been instrumental in introducing metal technology", and their arrival "coincided with major changes in the structure of primitive society". Such ambivalence left me wondering at the end whether the author had experienced a late

change of heart about "Beaker Folk", and been forced to take a scalpel to his text; or perhaps it is simply that he, like most of us, finds it difficult to adjust to a world without "Beaker Folk".

In many ways, Harrison engineers his own downfall. On p.13 he complains that "One bad habit which many archaeologists have had is to abstract beakers from their local settings and study them as something alien and isolated", which is exactly what has bedevilled study of this period in Europe, and is exactly the mistake this book perpetuates. For sadly, Harrison is largely concerned with his title, *The Beaker Folk*, and hardly at all with his subtitle, *Copper Age Archaeology in Europe*. But until the former is strictly subordinated to the latter, understanding of this exciting phase of European prehistory will advance very little.

There is a fundamental controversy about Beakers at the moment, and it should therefore be a priority in a work such as this to begin with a fair statement of the issues. As it is, for much of the book — until the last chapter in fact — Harrison skates over the difficult question of Beaker interpretations with brief, throw-away allusions, and in so doing is hardly fair to those who have been trying to open up the whole question to review. For example, on p.11 he is decidedly disparaging about the suggestion that Beakers may have been central to a beer drinking cult, but on p.15 they may "have been status symbols . . . perhaps on account of a special drink", and on p.69 "Bell Beakers could also have derived their importance from their contents, such as beer, mead . . .". If he looks again at what S.J. Shennan and I wrote in *British Archaeological Reports* in 1976, he will see that our concern was not to set up alternative models to explain Beakers, but to show that there are explanations, other than ethnic ones, which will fit the evidence. Harrison, like some others, has seized on the evocative peyote/beer suggestion, but does not mention the equally relevant Butt Beaker analogy.

The author's preoccupation with the Beaker aspects of this phase of prehistory results in much of the book being a rather dreary regional catalogue of contexts and sites in which Beakers occur. How much more illuminating it would have been had he remembered his sub-title, and con-

sidered the local background. As it is the Beakers too often float in a vacuum. The reader will find it nearly impossible to grasp how they relate to regional cultural developments, for frequent allusions make it clear that Beakers were absorbed by local populations and adapted to suit local circumstances. Thus Myrhøj in Jutland is a "Bell Beaker settlement" (p.35), but only one house had Beakers. Local Late Neolithic pottery dominates, but where does it fit in? In Hungary, the Beaker is intrusive in local Vučedol contexts, while over the Eastern Beaker province as a whole "between 80% and 90% of all pottery found in Bell Beaker graves [sic] is *not actually of a beaker shape at all* [my italics], but small cups, jugs, bowls and jars". So why should these be "Bell Beaker graves" and Myrhøj a "Bell Beaker settlement"?

What is needed here, if ever we are to make sense of the Copper Age in Europe, is a radical change in thinking. These are *not* "Bell Beaker" graves and settlements, but the graves and settlements of whatever local cultural group is involved, which happen to incorporate Beakers and Beaker influence to varying degrees. Usually Harrison at least acknowledges a local cultural element, for example in southern France (pp.114–116), and especially in Italy (p.123), where recent discoveries offer a good chance of putting Beakers into proper perspective. Alas, in Britain and Ireland, Harrison opts for the traditional Beaker invaders. "Bell Beaker settlements", as at Newgrange and Knowth, are mentioned without making it clear that only a fraction of their pottery is classic Beaker; a lot of it is decidedly *sui generis*, and the rest local in character. For me, the treatment of Britain is particularly disappointing: the author fails to examine objections to a distinct Beaker physical type; he does not scrutinize critically the so-called Beaker innovations; and he appears not to appreciate the supreme importance of the Lanting and Van der Waals regional approach, not just for British Beakers, but for later British prehistory generally.

It would be churlish not to compliment the author on his choice of illustrations, which make clear the skill and patience which went into producing Beakers; and as a panoramic view of the variety of regional Beaker styles this book will be invaluable. But Harrison's aim is to "give an intelligible account of the Copper Age in Western Europe" (p.6) and he does not seriously attempt the task. It is a perilous time to write on the Beaker phenomenon, but suicidal not to have a clear idea in one's own mind beforehand of what Beakers are all about. I confess that I would not be so rash as entirely to dismiss the Wee Folk; but I have no such qualms about the "Beaker Folk". □

Colin Burgess is Staff Tutor in Archaeology, Department of Adult Education, University of Newcastle upon Tyne, and author of *The Age of Stonehenge* (Dent, 1980).

Mummies: new light on ancient societies

A. Rosalie David

An X-Ray Atlas of the Royal Mummies. Edited by James E. Harris and Edward F. Wente. Pp.403. ISBN 0-2263-1745-5. (Chicago University Press: 1980.) \$55, £33. *Mummies, Disease and Ancient Cultures.* Edited by Aidan and Eve Cockburn. Pp.360. ISBN 0-521-23020-9. (Cambridge University Press: 1980.) £25, \$45.95.

GRAFTON Elliot Smith, Armand Ruffer, and Alfred Lucas led the field in the application of their disciplines — anatomy, histology and chemistry — to the study of ancient Egyptian mummified remains in the early part of this century. Since then, there has been a continuing interest in the use of medical and scientific techniques to amplify historical and literary knowledge of family relationships, the existence of disease, living conditions and funerary practices in this ancient civilization. Recent technological advances have encouraged further multi-disciplinary research, and the study of mummies has provided the Egyptologist with new information.

The discovery of the Egyptian royal mummies of the New Kingdom period (c. 1675–950 BC) was an archaeological find of considerable significance. The pharaohs and their queens, buried with great ceremony in rock-cut tombs at Thebes in Upper Egypt, were soon disturbed by robbers; their tombs were plundered and their mummies stripped of precious amulets. The high-priests of the god Amon re-buried many of these royal mummies in two resting places — the shaft of a tomb near Deir el-Bahri and a side chamber in a King's tomb in the Valley of the Kings — in the vicinity of Thebes. Here, they remained undisturbed until the chance discovery of the two caches in AD 1871 and AD 1898.

The mummies were subsequently removed to the Cairo Museum, where they were unwrapped and investigated by archaeologists and anatomists in 1889 and in 1912, when G. Elliot Smith produced his definitive anatomical work entitled *The*

Royal Mummies. This included the examination of kings, queens, princes, princesses, priests and some unidentified individuals. However, having the opportunity to examine only one of these mummies with radiographic techniques, he stated that "... examination with the aid of X-rays would, no doubt, have provided much additional information and I hope this will be done at some future time ...".

This opportunity did not arise until 1967, when the University of Michigan, which had previously collaborated with Alexandria University in a radiographic survey of the skulls of an ancient Nubian people, was invited by the Egyptian authorities to X-ray the royal mummies in the Cairo Museum.

Their aims were to add to historical and literary sources relating to the chronology of this period and to the lengths of the rulers' reigns; to assess the health of each individual; to check the reality of the royal sculpture and art forms against the physical evidence of the mummies; and, using a portable X-ray cephalometer, to determine the craniofacial variations and the degree of physical heterogeneity evident in the pharaohs and queens of the New Kingdom.

The *Atlas* provides the long-awaited reference work (a popular interim account by J. E. Harris and K. R. Weeks, *X-Raying the Pharaohs*, was published in 1973 by Charles Scribner's Sons). It does not give a detailed description of each mummy, but, in addition to the illustrative material, uses a series of self-contained chapters to highlight features of this unique collection. The chapters reflect the individual interests of the multi-disciplinary team, and include techniques of mummification, the practice of medicine and dentistry in ancient Egypt, evidence of physical and dental disease in the royal mummies, radiological techniques and findings, the genealogy of the Royal Family and craniofacial variation, and the age at death of the New Kingdom pharaohs (here, Egyptological and



Mummy of King Ramesses II, nineteenth dynasty.

From *X-Ray Atlas of the Royal Mummies*

physical evidence, sometimes at variance, is assessed separately). Such an approach, as admitted by the editors, results in a degree of repetition and an overlap of material in some instances. However, it is useful and interesting to see the physical and historical evidence summarized independently.

Another difficulty which the investigators faced, because of the paramount consideration for the safety of the royal mummies, was the necessity of examining the mummies in their modern wooden cases in the museum, instead of under near-ideal conditions in a modern radiology department where sophisticated equipment and a variety of radiological techniques would have been available. There is a consequent lack of quality in the X-rays and possibly in the completeness of the evidence obtained.

However, some of the most interesting results included the observation that, unlike certain other Egyptian mummy collections, these royal mummies showed a marked degree of cultural and genetic heterogeneity. This reflected the presence of foreign princes and princesses in the Egyptian royal family of the New Kingdom rather than the averred custom of royal consanguineous marriages undertaken to ensure succession to the throne. Another tentative conclusion, reached through the use of scientific techniques, was that a mummy known previously as the "Elder Lady" was in fact probably Queen Tiye, a close relative of Tutankhamun.

Nevertheless, the caution which historians should observe in utilizing the results of such scientific investigations is well illustrated by one of the conclusions drawn here. This concerns the proposed relationship between the pharaohs Amenhotep III and Tutankhamun. Previous investigations of the mummies attributed to these kings, undertaken by Professor R. G. Harrison and his colleagues, and based on palaeoserological and anatomical observations, had indicated that Amenhotep III was probably Tutankhamun's father. However, a different conclusion is drawn here: "The cephalometric X-ray records would not support the . . . contention. It is most improbable that the mummy designated Amenhotep III was the biologic father of King Tutankhamun".

The second book reviewed here, *Mummies, Disease and Ancient Cultures*, deals with a wider range of techniques and also shows how a variety of societies have preserved their dead, either unintentionally or by artificial methods, as "mummies". This study again shows the value of a multi-disciplinary approach. In the long term, it is hoped that the work will go some way towards answering questions concerned with the origin of diseases and their evolution in different societies. To this end, the team is committed to a detailed pathological study of world-wide collections of mummies.

A summary is given of the work to date, and although Egyptian and South American mummies have received most detailed attention at present, a survey of other mummified remains is included — those discovered in the United States, the Aleutian Islands, Alaska, the bog bodies of Denmark, mummies in Australia and Melanesia, and Japanese examples.

A complete section is devoted to reports of the techniques involved in this study which include physical anthropology, radiology, histology, electron microscopy, palaeoserology and palaeobiochemistry. Not only does the book illustrate the various methods — heat-drying, freezing or the application of dehydrating agents —

by which a body can be preserved, but it also shows how diverse are the religious and environmental reasons which bring about mummification in different parts of the world.

Both books reflect a resurgent interest in the application of scientific techniques to the physical remains of ancient societies, which, used with due caution and appreciation of the possible fallibility and limitations of such investigations, can provide the historian with a valuable additional insight.

Rosalie David is the Egyptologist at the Manchester Museum, University of Manchester, and Director of the Manchester Egyptian Mummy Research Project.

The cradle of geology refurbished

John Sutton

The Geology of Europe. By Derek V. Ager. Pp.560. ISBN 0-7084-115-2. (McGraw-Hill: 1980.) £16, \$38.40.

IN MY opinion this is much the best account of the geology of Europe to have appeared since Gignoux published his *Géologie Stratigraphique* over 50 years ago. Derek Ager describes his subject as Europe seen by a stratigrapher-palaeontologist. He treats his material very differently from Gignoux and the contrast is illuminating. Gignoux drew nine-tenths of his examples from Europe, and, in so doing, produced a comprehensive account of its geology which started with the oldest rocks, proceeded through younger formations to the most recent, providing abundant stratigraphical successions and the like on the way. Ager throws this approach out of the window; there is hardly a stratigraphical table in the book, so that he has space for much else.

Ager divides Europe into four regions on the basis of the relative ages of the last extensive period of metamorphism and deformation to have affected them. The geological history of each region is then described from start to finish — a method which brings out relations of older and younger events particularly well. Each of the four sections ends with a concluding chapter a few pages long. I was particularly struck with Ager's concise, five-page summary of Precambrian Europe. Although, as he puts it, "the Precambrian history of our continent is for the most part lost in the mists of metamorphism", he packs a telling account into that brief chapter.

The main virtue of this book is the way in which it brings together details to give a geological picture on a continental scale. Derek Ager has an interesting approach to geophysics; he appreciates "the hurricane of fresh air that new geophysical thoughts breathed into our subject". He goes on to write what is one of the best accounts of

any continent, set out on precisely the scale which would make it possible to compare the geological history of basins and uplifted crystalline blocks with the results of geophysical surveys. Such comparison, however, he leaves to others. He gives a good defence of his treatment, and in his introduction makes the point that any additional matter has to be balanced by the omission of something else. However, in a second edition I would hope at least to see a discussion of outline gravity and magnetic maps, and an account of crustal thickness.

Of all the continents, Europe is probably the most difficult to handle. Its geology is discussed in 20 languages, it is criss-crossed by frontiers and its complexities are probably the most apparent of any continent. Ager sails through all these difficulties triumphantly. He has travelled the continent repeatedly, he lists friends and helpers in every country; he is bold in his summaries and he writes with great clarity, and is not afraid to say what he thinks is important. Personal touches abound; we hear about the view from his study window and the nature of the rock surmounting the Ager garden gate. Historical figures pop up from time to time: Drake's game of bowls and the sailing of the Pilgrim Fathers illuminate the account of Plymouth Hoe. Whether, in consequence, one remembers which stage of the Devonian outcrops below the turf is probably a matter of individual experience. I expect every reader will find some of the allusions obscure; there are plenty of traps for the translators who will certainly get to grips with this most useful book. Unless they know their Sherlock Holmes and Irish mythology, we may yet see *chien de chasse des Basquervilles* and *Blarneystein* among the faunal lists and formation names of the future.

John Sutton is Pro Rector and Professor of Geology at Imperial College, University of London.

Clocks in grandfather's rocks

G.Y. Craig

The Abyss of Time: Changing Conceptions of the Earth's Antiquity after the Sixteenth Century. By Claude C. Albritton Jr. Pp.251. ISBN 0-87735-341-7. (Freeman, Cooper: 1980.) \$12.75 US only.

THE essence of this story is quite dramatic. The calculated age of the Earth has increased from a little less than 6,000 years to more than 4,500 million years in the last three centuries. The Earth, of course, has remained untouched by this spectacular revolution apart from the addition of these few hundred years. What *has* changed has been the consensus of opinion by philosophers, theologians, physicists, geologists and evolutionists as to how old the Earth actually is; and it is salutary to be reminded by younger colleagues that estimates of the age of the Earth have doubled even since I was a student in the 1940s. Most fortunately this precipitous rate of ageing seems to have stopped.

The two main ways of looking at the passage of time in geology are not dissimilar from the ways in which we recall our own lives. Events can be put in a rough sequence and then, often with difficulty, slotted into the calendar ("Remember that

The discovery of radioactivity at the end of the last century led to increasingly precise radiometric dating of those rocks containing suitable isotopes; and the two scales, relative and radiometric, can now be correlated to provide a chronological scale of historical geology. This scale is still being extended and refined. Metamorphic rocks in Greenland, for example, have recently been dated at 3,750 million years, the oldest rocks so far discovered.

Dr Albritton's account arose out of a series of lectures given to students and staff at various academic institutions in the United States. The slim, attractively bound text follows the traditional and by now well-worn path through the maze of geological time by examining the contributions of its protagonists. Leavened by portraits and anecdotes, it is easily read and clearly reveals the influence of religion, conceit, catastrophe, fanaticism, feuds, ignorance, intuition, rocks and radioactivity on the evolution of the concept of geological time.

In spite of jazzy chapter headings — Malta's cradle; Solids within solids; Mr Hook(e); A sacred theory of the earth; Telliamed's story; The epochs of nature; View from the brink; Father William; The four-dimensional jigsaw puzzle; A question of tempo; A plenitude of events; The great tree of life; Kelvin; A numbers game; Radiometric dating; World enough and time — the leading actors are readily unmasked. For the reader who does not care to play this game, the protagonists turn out to be Steno; Hooke; Burnet; De Maillet; Hutton; Smith; Werner and other early stratigraphers such as Sedgwick and Murchison; catastrophists and uniformitarians (I've now broken a New Year resolution not to use that word!) especially Cuvier and Lyell; observers of volcanoes and glaciers (Scrope and Agassiz); Darwin; and lastly late-nineteenth and twentieth century scientists. Pleasingly, the book includes an appraisal of Arthur Holmes' brilliant contributions to the construction of a geological time-scale. The last chapter contains a typical student-textbook analogy: by compressing the major events of the Earth into 365 days, primates appear on Christmas Day and *Homo sapiens* arrives a mere hour before the dawn of the New Year. Dr Albritton is obliged to devote part of this chapter to a rebuttal of creationist dogma, now so common in parts of the States. Such concern is understandable for he teaches in Texas where fundamentalist views are not unknown.

Yet somehow the excitement of what should be a gripping story is not there. In one form or another it has all been done before, albeit with the historical anecdotes cut to a minimum. Last year two long articles by John MacPhee in, of all places, *The New Yorker*, conveyed a similar

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

William Thomson, Baron Kelvin (1824–1907).
Detail from a cartoon in *Vanity Fair*.

summer when we went to Brittany and Aunt Doris got . . . Now what year was that?"). And so it is with rocks. Geologists during the nineteenth century pieced together a sequence of geological events which enabled them to construct a relative time scale. Concurrently various attempts were made to establish when these events occurred and especially how old the Earth might be. Such calculations involved rates of erosion and deposition, the rate at which salt is being added to an already salt sea and the rate at which the Earth appeared to be cooling.

Mary Evans

SCIENTIFIC BOOKSHOP

H.K. LEWIS can supply works in all branches of Pure and Applied Science. Catalogues on request. Please state interests.

SCIENTIFIC LENDING LIBRARY

Annual Subscription from £7.50.
(Available in U.K. only)

Reduced rates for multiple subscriptions.

Prospectus post free on request.

Quarterly List of New Books and new editions added to the Library sent post free to subscribers regularly.

**H.K. LEWIS
& Co. Ltd.**

136 GOWER STREET,
LONDON, WC1E 6BS

Telephone: 01-387 4282
Telegrams: "Publicavit,
London, WC1E 6BS."

Circle No.19 on Reader Enquiry Card.

Bring Your Library Up-to-Date With This ANS Edition ... Nuclear Energy in Germany

Winnacker/Wirtz

6 x 9 hardbound edition
\$37.00

**American Nuclear Society
555 North Kensington Avenue
La Grange Park, IL 60525 USA**

Please take my order for _____ books.

☐ Payment in full enclosed.
☐ Bill me and I accept postage
and handling charges.

Name _____

Org. _____

Street _____

City _____

State _____ Zip _____

Circle No.02 on Reader Enquiry Card.

message but in a more entertaining style, perhaps because his serious but flowing story was laced with irrelevant cartoons and adverts including fast cars, whisky, cameras, jewellery and beautiful women, but more probably because he linked history with current field-work in different parts of the States.

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

James Ussher, Bishop of Armagh — calculated the date of creation as 4004 BC.

So here I am, vaguely dissatisfied with a perfectly adequate book simply because of a feeling of *déjà vu*. How can one be constructive? More might have been made of Archbishop Ussher (1581–1656) and his method of calculating the date of creation (4004 BC) and such subsequent events as the grounding of the Ark on Mount Ararat on May 5, 1491 BC (a Wednesday!). His chronology was incorporated in the most widely read book of the day — the King James VI bible — and so had an undue influence in Protestant countries. Many of the subsequent philosophical and scientific investigations described in the *Abyss of Time* were carried out in the shadow of Ussher's precise chronology. Astronomical calculations and coral clocks are not referred to but they show that the Earth is now rotating more slowly than it did during the Devonian when a year was all of 399 days. Above all, data from meteorites and Moon rocks reveal fascinating events earlier than any now preserved on Earth and provide independent factual evidence of the great antiquity of our Solar System and planet. They could have been the subject of a separate chapter.

And what of the future? The abyss is just as much in front of us as behind. The Earth is steadily losing its heat. As it continues to cool the lithospheric plates will thicken and stop moving: eventually there will be no more volcanic activity and no more earthquakes, but the land will continue to be eroded seawards, there to remain because uplift will no longer occur. The rock cycle will finally have come to an end: but there is perhaps consolation in the fact that the Sun will continue to shine until in another 5,000 million years it too becomes cold. And after that? But Albritton does not indulge in such fantasies. □

G. Y. Craig is Professor of Geology at the University of Edinburgh.

Trap for the unwary

William H. McNeill

The Fates of Nations: A Biological Theory of History. By Paul Colinvaux. Pp.383. ISBN 0-671-25204-6. (Simon and Schuster: 1980.) \$12.95 US only.

For the intellectual wanderer, impatient with the myopia of specialization, there are great risks in invading an unfamiliar turf. Jealous experts lie in wait for the newcomer to display his ignorance of their discipline by some slip of tongue or pen, and the complications of the new subject may not conform readily to the insights the newcomer brings from some other discipline.

This, unfortunately, is the case with Paul Colinvaux's book, *The Fates of Nations*. Colinvaux is an ecologist whose speciality is Ice Age Alaska and the land bridge which once allowed animals and plants to migrate from Asia to America across what is now the Bering Strait. For general readers, however, his best-known work is a summation of recent ecological niche theory which goes by the catchy title *Why Big Fierce Animals are Rare* (recently issued in paperback by Penguin). As an historian, who, like Colinvaux, has found it irresistible to wander across recognized disciplinary boundaries, I found this essay instructive and persuasive. But now when he takes on my own field of expertise, the human past, and seeks to illumine it by bringing ecological niche theory to bear, I find Colinvaux's ideas inadequate and his information radically defective.

As he engagingly explains at the end of the book in a "Note on Sources and Substance", he began to think about a "formal theory of history" in 1971, relying on the history "which is the common knowledge of Christendom", i.e. on what he remembered from his school days in England. This he reinforced by reading Toynbee's "overview of history's facts", and by consulting a few experts on battles — Livy, Tacitus, Arrian — as well as more modern writers like Liddell Hart and J. F. C. Fuller. He found the barrister Edward Creasy's *Fifteen Decisive Battles of the World*, first published in 1854, especially useful, while Field Marshal Montgomery's *A History of Warfare* supplied him with "the clearest insight into what gave victory in many a crucial battle". Thus, as Colinvaux tells us himself, his method of mastering history was studiously to disregard what professional historians have been doing since World War I to modify the patriotic and political history so ably caricatured in *1066 and All That*.

With *1066 and All That* plus school-boy Latin as his guide, and Toynbee as guarantor of "facts", what does an intrepid ecologist discover in the human past? The answer is a rhythm of rise and fall, predictable in the light of ecological niche theory. It may be summarized as

follows. Initially, history starts when a group of human beings develops a new technique that allows them to enlarge the ecological niches in which they live. This leads to a rise in population, for human beings have always pursued a breeding strategy — the large young gambit — that requires each breeding couple to raise as many children as they think they can afford. This strategy, Colinvaux assures us, has never altered, though he does not assert that it is instinctive.

Once enlarged niche spaces start to fill up as population increases, the prospering community faces a crisis. Since growing numbers with no increase in resources will force them to accept narrower niches again, they must look abroad to find the means to sustain their heightened expectations. Trade, colonization and aggressive war, aimed at taking land away from those in possession, are the alternatives such a population faces, and this is what history is all about. Successful communities expand and create empires; but sooner or later, expansion ceases. Colinvaux is completely silent on what stops expansion. On his principles I can see no reason for the process he sketches not to continue indefinitely, until the most successful aggressor has created an empire of the Earth. But the "common knowledge of Christendom" makes clear that a limit was set somehow, and Colinvaux shows his deference to facts by stopping the expansion of the Roman Empire at the Rhine and Danube — though, as I say, he gives no reason for this interruption of a process which he has declared to be both predictable and inevitable.

Once territorial expansion stops, the continued growth of population requires narrowed niches; and narrowed niches will be accepted only reluctantly by the population concerned. This means an oppressive government — a police state in fact, in which the rich, clinging to their broad niches, impose their power on the poor by dint of "regimentation, bureaucracy, class, rationing and caste" (p. 64). Such a regime is vulnerable from without, and eventually barbarian invaders, technologically abreast of the latest methods of civilized warfare, will succeed in destroying the oppressive regime by cutting off food supplies to the cities. The newcomers institute a barbarian pattern of life in which local consumers depend on local food production. This sharply reduces the total number of inhabitants, since the urban masses, congregated under the regime of civilization into vast cities, are unable to maintain the breeding strategy that had hitherto led to an ever-mounting population. They simply die off, leaving a sparsely settled countryside to start the cycle of civilization over again.

Echoes of Toynbee's cycle are obvious, though the emotive tone that Colinvaux's ecological vision imposes is utterly at odds with Toynbee's humanistic and religious values. There are, too, some serious anomalies in this thesis, for the modern cycle of civilization does not conform closely to the classical Mediterranean pattern.

An even greater anomaly is introduced in Colinvaux's final chapter "The Shape of Things to Come". After exploring the implications of this theory, which predicts that growing populations will launch aggressive war in order to maintain enlarged ecological niches, he apparently reverses himself by saying that it "may well be that our breeding strategy will yield even smaller families for the moderately affluent than they have now" (p. 313), since one may hope that "new roles for women will hold down the population and let us all live in freedom" (p. 314). So because "we breed in human and controllable ways" and can understand what is happening to us, we "need fear neither our future nor our fate" (p. 321). But if human breeding strategy is alterable now, why not in the past? And if one group of people in this twentieth century does decide to have fewer children in order to maintain broader niches, what prevents others from breeding incontinently and filling in the space occupied by the rich and continent, not by conquest, but by the less spectacular method of in-migration?

This sort of population shift is occurring before our eyes, as *Gastarbeiter* in Germany, illegal aliens in the USA and Asians in England attest. It occurred in distant ages as well. The replacement of Sumerian speakers by Semites in ancient Mesopotamia in the 3rd millennium BC and the bitter jibes of the Roman poet Juvenal against the Greeklings whom he saw taking over favoured places in Roman society, are only two examples of the past importance of peaceful migration. Indeed, all great imperial cities of times past have been polyethnic in their population. They were places where diverse peoples congregated and died off due to the intensified disease exposure involved in urban living before the days of modern sanitation. Colinvaux seems unaware of this traditional population dynamic, well attested though it is for early modern Europe when statistical records begin to become available to historians. Indeed he is ignorant of all historical demography, though for a theory that puts such emphasis on population growth as the fundamental disturber of history he surely ought to have guarded himself against the sort of reproach I make of his procedures by at least consulting a few elementary textbooks on historical demography. If he had done so, he would have discovered that historians are agreed that the population of the Roman Empire declined sharply in the second and third centuries AD, recovered partially in the fourth century and then plummeted again in the fifth. Colinvaux's

picture of an ever-rising population pressing harder and harder on available niches therefore runs counter to all modern historical scholarship.

As a matter of fact, Colinvaux is aware of some of the signs of depopulation that manifested themselves in late Roman times. But he dismisses empty fields and unoccupied border regions where Roman authorities settled barbarian tribesmen from beyond the frontier as a predictable consequence of population growth! The argument is sufficiently bizarre to deserve quotation in full:

From the time of troubles onward the population shifts within the state. There is a drift from the land as peasants are displaced in the interests of increased production. This happened in ancient Greece and Rome no less than in the time of the enclosures in Tudor England, or in modern industrial states. The process can be seen in the development of every civilization. Feeding great numbers of people is more easily done if they are brought together in dense settlements; running the agriculture needed to supply these dense settlements is more efficient in the larger agricultural units of agribusiness. And the "drift from the land" is a predictable consequence of a dense population growing denser [pp. 83-84].

To test this hypothesis one might look for some indication of the growth of cities while fields emptied out. Roman cities from the 3rd century onward conveniently provided such an index by erecting ring walls necessitated by the declining level of public security in the empire. By modern standards these enclosures were tiny indeed, and diminish with each rebuilding. Where is the massive, ever-growing urban population Colinvaux's theory requires? Even on a purely theoretical level, and without the vulgar appeal to mere data, one would suppose that before propounding his ridiculous hypothesis Colinvaux might have asked himself how it could be more efficient to carry food from the fields to distant urban consumers (in ox carts presumably), and how "agribusiness" could plough and reap in the absence of modern machinery, and without numerous labourers on the spot to do the work. The fact is that human populations rise and fall in complicated and only partially understood ways, and in response to a far wider range of shifting conditions than Colinvaux's system admits.

My judgment of *The Fate of Nations* is that in venturing beyond the boundaries of his discipline Colinvaux has come a cropper. This is a pity, for surely there are useful things historians might learn from ecologists, and an ecologist who took the pains to acquaint himself with historical literature might indeed be in a position to shed new light on familiar historical conundrums.

William H. McNeill is Professor of History at the University of Chicago, and author of The Human Condition (Princeton University Press, 1980).

THE EARTH'S PROBLEM CLIMATES

Second Edition
by Glenn T Trewartha

In this new and extensively revised edition, the author takes into account new information and climatic data that have become available since the original edition appeared in 1961.

Trewartha offers an inventory of the more important of the earth's problem climates, brief and concise descriptions of their characteristics, and suggested explanations for at least some of them; in so doing, he draws together the results of numerous definitive studies in dynamic climatology.

Wisconsin, May 1981, 352 pp, illus, cloth £13.50

VOLCANOE OF THE EARTH

by Fred M Bullard

"... a first-rate and up-to-date account of one of the world's most intriguing phenomena."

-*Explorers Journal*

Texas, 597 pp, 24 colour plates, 183 b&w illus, now in paperback at £12.00

AMERICAN UNIVERSITY
PUBLISHERS GROUP LTD

1 Gower Street
London WC1
Telephone (01) 580 3994

Circle No.41 on Reader Enquiry Card.

SCIENCE AT THE WHITE HOUSE

A Political Liability

EDWARD J. BURGER, JR.
foreword by DON K. PRICE

From the day the first atomic bomb was dropped, scientists gained a new image in the public's eye. In the succeeding decades, there seemed to be no limit to what technology could accomplish, from military weapons to moon walks. *Science at the White House* looks at what happened when America's new experts were called in to serve as advisers to the president.

As cases in point, Burger assesses the government's efforts at planning for key issues that emerged during his tenure as a science adviser: issues concerning national health care, environmental regulation, and population and family planning, among others. 208 pp, £9.00.

THE JOHNS HOPKINS
UNIVERSITY PRESS

Ely House, 37 Dover Street,
London W.1.

Circle No.47 on Reader Enquiry Card.

One hundred years of the Natural History Museum

D.T. Donovan

Alfred Waterhouse and the Natural History Museum. By Mark Girouard. Pp.64. ISBN 0-300-02578-5. (Yale University Press: 1981.) Hbk \$12.95, £4.95.



Alfred Waterhouse, 1830–1905, from a portrait by Sir William Orchardson.

THE Natural History Museum is the most splendid of the South Kensington museums; it has a vigour and coherence which is a tribute to the architect Alfred Waterhouse. So it comes as a surprise to learn that the building has a complicated history, and was nominally designed by somebody else.

The natural history collections were for long kept in Bloomsbury with the books and antiquities of the national museum. By the early nineteenth century antiquities were being added in large quantities, and by mid-century many thousands of natural history specimens were accumulating every year. The new British Museum building, begun by Robert Smirke in about 1823, was already too small by the time it was completed in about 1850.

Mark Girouard begins his account with

Richard Owen's vision of a great new museum in 1858, which was the origin of the present building. Removal of the natural history departments from Bloomsbury had been considered for some time, though a practical solution was not found until after 1851 with the development of South Kensington as a centre for education in the arts and sciences.

Owen, already famous for his contributions to comparative anatomy, had been appointed Superintendent of the Natural History Departments in 1856. In the same year, Antonio Panizzi, Keeper of Printed Books, was made Principal Librarian (i.e. Director) of the museum. Panizzi had long wanted to get rid of the natural history departments, and said so on every possible occasion. Yet even with Owen and Panizzi working towards the same end, it was another 25 years before the natural history collections were actually moved out of the Bloomsbury building. Initially, at least, almost the whole of the scientific establishment was opposed to such a move, and succeeded in influencing Parliament against it for some time.

In 1863, however, the Government bought the present site. On it stood the exhibition building of 1862, built by the official architect Francis Fowke, and widely disliked. Parliament refused to purchase this building, which was demolished, and an architectural competition for a great new museum complex was held in 1864. It was won, to everyone's surprise, by Francis Fowke, but he died before his plan could be executed.

Alfred Waterhouse, newly arrived in London from successes in Manchester, was appointed to carry out Fowke's design, but it was not until 1873 that building was started. By this time only the Natural History Museum — about one-quarter of Fowke's complex — was actually going to be built, so the building had to be drastically re-designed. Girouard traces the

influence of Fowke's design by comparing successive versions of the building. However, there is little perceptible resemblance between Fowke's drawings and the present building, and Waterhouse's standing as the architect is fully justified.

The plan of the museum, however, is neither Fowke's nor Waterhouse's; it stems from a sketch by Owen in 1859. This shows main east–west galleries with numerous galleries opening off them at right angles, the arrangement in the present-day museum.

Owen had always wanted his museum to be enriched by appropriate decoration. He and Waterhouse devised a scheme in which representations of living animals and plants adorned the west wing, intended for zoological collections, and extinct ones (restorations, not fossils) the east wing, where fossils and minerals were to be housed. Waterhouse's drawings were translated into models and cast in terracotta; comparative pictures in the book show how faithfully this was done.

The book has been sponsored by the Museum as part of the celebration of the centenary of its opening in 1881. This centenary is, unhappily, marked by controversy over the current proposal to demolish the rear single-storey galleries, thus destroying Owen's original plan.

Mark Girouard has written an attractive and interesting book which explains the history of the building and its architectural origins. It is well illustrated with pictures of the museum as it is, Waterhouse's working drawings and earlier versions of the design. Only the photographs of the interior are disappointing and do less than justice to the building. Otherwise, the book makes a handsome souvenir of this great embodiment of Victorian scientific achievement. □

D.T. Donovan is Professor and Head of the Department of Geology at University College, University of London.



As it might have been — the winning, but unexecuted, design submitted by Francis Fowke in 1864. On the death of Fowke, Alfred Waterhouse was appointed as architect and the museum was built largely to his designs.

Intellectual cross-fertilization and evolution

A.J. Cain

The Evolutionary Synthesis: Perspectives on the Unification of Biology. Edited by Ernst Mayr and William B. Provine. Pp.487. ISBN 0-674-27225-0. (Harvard University Press: 1980.) \$25, £15.

THIS book looks out upon an incredibly rich panorama of personalities, groups, institutions, countries, attitudes, theories, arguments, experiments, facts, false facts (even a very few artefacts), misunderstandings, barriers emotional, linguistic, dogmatic and intellectual, all adding up to one of the greatest intellectual advances of our time — the unification of different biological sciences by a single neo-Darwinian theory of evolution in the first half of this century. Some of the principal actors have contributed invaluable accounts of their own roles: among them are Mayr, Simpson, Dobzhansky, E.B. Ford, C.D. Darlington, Curt Stern; and there are extremely valuable surveys and analyses by biologists, historians and philosophers (Lewontin, Gould, Mark Adams, Dudley Shapere etc.). The book itself is now history and, since the 1900–1940 period qualifies as history (not merely as out-of-date), it will inspire and provoke an ebullition of PhD theses, papers and books worse than the great Wittgenstein industry. One of the virtues of the book is that while making thoroughly clear the limitation of science by the inadequacies of personalities, institutions and governments, it does present beautifully, especially in genetics and cytology, the intellectual impact of actual scientific discoveries and their role in changing people's thinking.

The book is organized by subjects (Part I, cytology, embryology, systematics, botany and so on) and by countries (Part II, Russia, Germany, England, France, USA), with a concluding survey (no, I will not say "overview") of the whole subject and a miscellany of very useful biographical contributions, rather over-titled as "Biographical Essays".

Especially valuable to those with little German and less Russian is the survey by countries, and a saddening spectacle it presents; the superb beginning in Russia, based on an excellent naturalists' tradition of population genetics, aided at first by the ideals of education and especially science during the Revolution, then brutally terminated as a result of the charlatan Lysenko's power-politics. France did not contribute at all to the synthesis except for Teissier's and L'Heritier's invention of the population cage and work on balanced polymorphisms. Both started as mathematicians (the probable reason for Teissier getting his chair at the Sorbonne is fascinatingly non-scientific), uncontaminated by the ideas of the French Establishment.

Both Boesiger and Limoges have some very hard things to say about the academic establishment in France, and, so far as I can see, they are thoroughly justified. Boesiger remarks "France today (1974) is a kind of living fossil in the rejection of modern evolutionary theories; about 95 percent of all biologists and philosophers are more or less opposed to Darwinism."

Limoges comments equally astringently on the language barrier, the closed university system ("no one can hold a chair in a French university who is not French and who does not have a French doctorate") and the centralization of the system, providing for career control by a few senior professors in Paris. Although that has now been considerably modified, Boesiger emphasizes the centralized control of the research organizations. One can add that there is considerable pressure on French workers to publish in French. Certainly there are in France today some excellent and up-to-date evolutionists — one thinks immediately of Maxime Lamotte in Paris, and a botanical group at CNRS Montpellier, as well as Zuckerkandl and others at that university; nevertheless it seems to be true that a largely self-perpetuating Establishment, accepting only conformist successors, has made France, as far as evolutionary studies go, into an intellectual island. Yet the new recruitment every year of fresh minds is one of the most valuable processes in any intellectual society.

Germany is well treated by Mayr, Rensch and Hamburger. Mayr brings out important administrative limitations (not enough chairs and the nature of professional duties), Curt Stern the difficulties of institutions separated by distance (Berlin-Buch and Berlin-Dalhem!). Little is said about Nazi influence. Rensch, in an essay very sympathetic to his predecessors and colleagues, brings out many reasons for the slowness of the synthesis in Germany — apparently contradictory phenomena in genetics and palaeontology, speciation apparently concerned only with trivial characters, the apparent internal drive of the organism in development and regulation, independent of the environment, religious convictions and so on, which, however, operated elsewhere as well. The slowness of penetration of genetical results in Germany certainly needs study.

Hamburger brings out a further point (which Boesiger applies also to Cuénot in France): "the widespread German tendency to have a unified *Weltanschauung*, a more general overview of metaphysical and scientific creeds than just having *Weltanschauung* and scientific work compartmentalized side by side".

The history of the evolutionary synthesis in the USA is particularly well treated, and deservedly so, and there are many studies of that intriguing man T.H. Morgan and

Minerals from the Marine Environment

Sir Peter Kent, FRS

Resource and Environmental Sciences Series

A review of the full spectrum of mineral resources development as seen against the environmental background of the seas and oceans.

£3.95 paper 96 pages

The Ecology of Natural Resources

Second Edition

I.G. Simmons

The author has strengthened the new edition both by updating it statistically and by changing it conceptually. Increased emphasis is given to energy as a resource and as a link with other resource processes.

£8.50 paper 448 pages

Additions to the *Studies in Biology* series:

124 Light and Plant Life

J.M. Whatley and F.R. Whatley, FRS

£2.75 paper 96 pages

125 Lipids and Polysaccharides in Biology

Anna J. Furth

£2.10 paper 72 pages

126 Bioenergetics of Autotrophs and Heterotrophs

John W. Anderson

£1.95 paper 64 pages

127 Invertebrate Respiration

Rufus M.G. Wells

£2.25 paper 76 pages

128 Immunobiology

Christopher J. Inchley

£2.50 paper 80 pages

129 Growth Regulators in Crop Production

Leonard C. Luckwill

£1.95 paper 64 pages



Edward Arnold

41 Bedford Square,
London WC1B 3DQ

Circle No.15 on Reader Enquiry Card.

his close associates. The conversion of this school to a neo-Darwinian attitude by their own discoveries in genetics and karyology, powerfully aided by their openness to ideas and visitors from abroad, was, however, only partial in some cases because of the difficulties of extending genetical findings to macroevolutionary phenomena, including speciation. Mayr makes an especially important contribution in emphasizing that the synthesis *was* a synthesis; it was extremely difficult for many years to reconcile Darwinism and the new science of genetics, and it was systematists in particular who contributed the concepts of adaptive geographical variation, variation in whole populations under environmental selection, and partly adaptive speciation (a macroevolutionary process). Yet it must be added that it was largely the wholly unjustified insistence of many systematists that characters of species were non-adaptive that misled many of the earlier evolutionists into doubts about the efficacy of natural selection and the continuity of micro- and macroevolutionary phenomena.

Equally fascinating is the comparison between subjects. Botany (not the botanical part of genetics) was almost as poor a contributor among subjects as France among countries, and Ledyard Stebbins's article is as thought-provoking as Boesiger's or Limoges'.

When one stands back from it all, several major thoughts come to mind. The difficulties are highly instructive, but little more than one would expect in any intellectual controversy; indeed Simpson is probably right in indicating that given the human character, the synthesis was achieved about as fast as could be expected. But some topics are hardly touched on. I suspect (in a few instances, I know) that religious considerations came first, even if they were seldom expressed, in some scientists' minds, and conditioned heavily what they were prepared to accept scientifically. Under this heading I include humanism when it refuses the application of the synthesis to mankind, and of course Marxism with its leaning to some form of Lamarckism (only too well documented in this book).

Secondly, and allied to the preceding, the role of chance as an explanatory factor is hardly touched on — yet, on the one hand, as a means of explaining away inconvenient phenomena it has been invaluable as an alternative to natural selection; on the other, it has added to the revulsion of many people against natural selection that what the process has to work on is the effects of mere chance.

But lastly, there is almost no indication in the book that the recognition of natural selection as the major (or nearly exclusive) agent of evolution was ever due to any work on natural selection in the wild, except very indirectly. A simple-minded (or extremely intelligent) Martian might well think that if it was debated whether natural

selection operated in the wild, the thing to do was to go and look for it. Chetverikov might have done so if politicians had let him; Haldane thought it much too difficult. The whole debate over the efficiency of natural selection in the wild was conducted by argument over indirect evidence, and indeed to a large extent it still is.

The present book contains little material towards a history of actual field-work directly on natural selection, except for E.B. Ford's contribution — and perhaps rightly, since it seems not to have affected the synthesis. It does contain valuable hints towards a history of why no one has attempted it — for example, Lewontin's statement that "Wright's notion that random processes allow the exploration of the field of gene frequencies . . . frees the evolutionist from having to ask such

questions as why did some Ceratopsian dinosaurs have three horns and some have two and some have a frill and some not? ". Of course it does nothing of the sort except to those who want it to. It was not *shown* in the first place that the development of different ways of fighting was due to any random process, except on the presumptuous assertion "Because I can see no sense in this phenomenon, not having worked on it, therefore it must be random". The use of random processes in explanation even at the present day is rather like John Henry Newman's use of St Augustine's text to precipitate him into the Roman Catholic church (see Geoffrey Faber's *Oxford Apostles*) — using a bad reason for doing what we intend to anyway. □

A.J. Cain is Derby Professor of Zoology in the University of Liverpool.

Avian evolution transcends philately

Alec Panchen

The Age of Birds. By Alan Feduccia. Pp.208. ISBN 0-674-00975-4. (Harvard University Press: 1980.) £12, \$20.

OF ALL those who concern themselves with the palaeontology and evolution of vertebrate animals, it is arguable that bird palaeontologists have taken on the most difficult task. The anatomical uniformity of the majority of bird species is such that the identification of their individual fossil bones varies between difficult and impossible; further, there are numbers of cases (some cited by Feduccia) in which the avian status of particular specimens is still a matter for debate. It is probably for these reasons that while popular books on dinosaurs are legion, a book on bird palaeontology is unusual despite the innumerable amateur ornithologists that are the mainstay of almost every natural history society. One would hope, therefore, that there was a market for a popular book like Feduccia's as evidence that at least some of these amateurs have interests extending beyond the prevalent stamp-collecting approach to bird watching.

On the whole Feduccia serves his potential audience well. However, his opening chapter tends to hop about its subject matter in a manner reminiscent of small birds themselves, starting with a scene-setting paragraph on *Archaeopteryx* — geologically the oldest as well as the most famous of all bird fossils — but then switching to bird anatomy, "Darwin's finches", adaptive radiation, the concepts of homology and analogy, the reconstruction of phylogeny and the geological time scale, before settling on an account of the several finds of *Archaeopteryx*, with useful photographs of each specimen.

In his second chapter he gives an account

of theories of the evolutionary origin of birds, a fiercely debated subject still. When dealing with controversial issues Feduccia is usually sensible and it is not his fault that at least one of the principal protagonists in this debate has changed sides since the chapter was written. Following chapters deal with the evolution of flight and then with the fossil history (such as it is) and evolution of all the major groups of birds.

Feduccia has been in the forefront of some of the phylogenetic work he describes and much of the book is authoritative. However, particularly in the early chapters, there are some things to take exception to. The layman's naive view of the course of evolution, as a highway from simple organisms to human beings with less important byways branching off, will be reinforced by statements such as "ornithischians [a dinosaur group] were highly specialised herbivores and far too removed from the main line to have given rise to any other major group". It will be similarly reinforced by the use of "Age of Reptiles", "Age of Mammals and Birds" and so on, as in "The cotylosaurs [the earliest reptiles in the author's (outmoded) usage] appeared long before the Age of Reptiles . . .". There are also a few obvious errors — how, for example, can the two views of the original *Archaeopteryx* feather, pictures on p.12, be "main slab" and "counterslab" when they are not mirror images of one another? Also, the patagium (flying membrane) of the living flying lizard genus *Draco* does not stretch "between the forelimb and hindlimb" (p.40): it is supported by elongate ribs alone.

On the whole, though, the book is easy and pleasant to read. The style is clear, concise and interesting, albeit with some infelicities — "Members of this classifica-

tion . . ." (i.e. taxon); "... third alternative . . ." — and the chapters are short. The bibliography is excellent; here Feduccia refuses to "talk down" to the reader and quotes every reference, not just "further reading". However, the illustrations although numerous are often disappointing. They include black and white photographs of specimens and localities, and a number of frankly mediocre original drawings by several hands, the low standard of which is emphasized by some of the splendid and



Reconstruction of *Archaeopteryx* by Rudolf Freund.

famous restorations by Heilmann, Charles Knight and others with which they appear. Many of these cry out for their original colours. Presumably the decision to exclude the use of colour was the publisher's rather than the author's; nevertheless I imagine that it will affect sales adversely. This is a pity as the book is otherwise well produced and certainly deserves an audience. □

Alec Panchen, a vertebrate palaeontologist, is Reader in Vertebrate Zoology at the University of Newcastle upon Tyne.

Gorillas on a tightrope

Alison Jolly

The Natural History of the Gorilla. By A. F. Dixon. Pp.202. ISBN 0-297-77895-1. (Weidenfeld & Nicolson: 1981.) £16.50. To be published in the USA by Columbia University Press.

IF YOU want to know about gorillas, you need this book. Alan Dixon has summarized the present state of knowledge about all aspects of gorillas' biology from anatomy to conservation. Furthermore, he writes in a lucid, factual style which presents the information clearly to anyone from fifth form to professorial status, without either talking down or jazzing up.

The drawings (by Dixon himself) add both clarity and charm.

The book has several strengths. First, it deals equally with the results of laboratory work and results from the field. Dixon's own studies range from laboratory analysis of primate hormones and sexual behaviour, through the testicular atrophy of several adult male silverbacks in zoos (including Guy at London), to making field censuses of gorilla populations. He is thus well placed to appreciate both sides of the story. A second strength is the balance achieved between generalization and data from individual studies. In the chapter "How Close to Man?", he manages to explain quickly many of the bases of primate classification, sum up the evidence which suggests man's close relationship to gorilla and chimpanzee and the more distant relationship of these three African species to orang-utans, and yet include the divergent studies which indicate Asiatic, rather than African, origins for man. Similarly, when he describes wild populations, he includes data showing differences between the three gorilla subspecies, and even between western and eastern Virunga Volcanoes, adding appropriate caution about sample size. In a popular book, it is unusual to avoid generalizations about "the gorilla" meaning ten or twenty individuals on Mount Visoke.

Finally, Dixon covers the whole scope of the literature from Hanno (fifth century BC) up to 1980. This enables him to deal fairly with such experiments as Yerkes' on Congo (1926–1928), which is still one of the most complete accounts of a gorilla's problem-solving, as well as Redshaw's (1978) splendid Piagetian study of the development of infant gorilla intelligence. The conservation account is also up to date, to the death of Digit and others who had their heads and hands chopped off for tourist souvenirs.

The main reservation I have is that Dixon does not pursue his synthesis of detail quite far enough for my taste. I looked for a discussion of the relation of gorilla leaf-eating to metabolism, ranging and social structure, and comparison with the frugivorous orang and chimp. All the pieces are there, but Dixon does not explicitly link them up; similarly, reproductive strategy is split into several different chapters. This is clearly his choice — he gives us the data without fashionable theorizing, but he is better placed than most of us to deal with such theories.

I could also do with slightly more drama in the final chapter on conservation — perhaps some of Dian Fossey's emotion over Digit, or Alan Goodall's encounters with poachers in *The Wandering Gorilla* (Collins, 1979). Instead, Dixon has chosen restraint. Perhaps he is right that estimates of the remaining wild populations and their possible or likely extinction are even more horrifying.

The whole question of how to balance

professional against popular appeal bedevils primate studies. In the circumstances Alan Dixon's summary of the natural history of the gorilla for such a wide audience is laudable. Scholarly balance may seem faint praise until you add that it is balance on a very thin tightrope indeed. □

Alison Jolly is temporarily in the School of Cultural and Community Studies, University of Sussex.

Lure of the shore

J.T. Enright

Intertidal Invertebrates of California. By Robert H. Morris, Donald P. Abbott and Eugene C. Haderlie with 31 contributors. Pp.920. ISBN 0-8047-1045-7. (Stanford University Press: 1980.) \$45 US only.

This is a book for those who find the California shoreline at low tide a place of beauty, wonder and interest, and who want information on the common animals that live there. [Preface].

THE west coast of North America has one of the most diverse temperate-latitude intertidal faunas in the world. In California that fauna provides research material for dozens of marine laboratories and learning material for hundreds of high schools, colleges and universities. *Intertidal Invertebrates of California* is aimed at that limited — some might say, provincial — audience, but it could conceivably also lure naturalists from the entire world to view and study this showplace of diversity. Whether or not it eventually has that sort of incidental success, there is no question about its value and importance for the intended audience: every biologist, professional or amateur, who ventures into the California intertidal zone for the first — or the ten-thousandth — time should have this volume available.

Thanks to a grant from The David and Lucille Packard Foundation, the price is such that most users will be able to purchase their own copies. Even a first glance at the 200 photographic plates — all but three of them in colour, most with four to six separate portraits of the animals — will persuade the prospective buyer that this book is a bargain. The majority of the photographs were taken by Morris, but nearly 50 other naturalists contributed to the handsome illustrations. This portion of the book will be particularly treasured by all who find the "shoreline . . . a place of beauty [and] wonder".

With the photographs as the bait, Abbott and Haderlie, together with 31 other contributors, have provided nearly 700 pages of text, bibliography and index for those "who want information". Some 750 species are treated in detail in the 24 chapters. For each of entry, the text provides the classification of the animal; a morphological description, including

typical size and the features which distinguish that species from those with which it might be confused; and notes on habitat and geographical distribution, often with qualitative comments on abundance. Usually there is additional information on what is known of the animal's natural history. Each entry closes with literature citations, sometimes only one or two, in other cases several dozen. Detailed references are then provided at the end of each chapter.

The introduction suggests that this book was intended as a primary guide for identifying the animals, but this emphasis is, I think, misplaced; it will simply not do as a field manual. Its unusual format and sheer weight (about six pounds) militate against carrying the book along to the shore. Beyond the problems of getting book and animal together (since the reader is, rightly, urged not to bring specimens home with him), there are none of the traditional "keys" for identification; and the coverage is intentionally restricted to the more common and conspicuous elements of the fauna, meaning that using the plates and text for primary identification would be not only time-consuming but often misleading.

For field identification, then, one should instead take along Smith and Carlton's revision of *Light's Manual* (University of California Press, 1980). Thereafter, *Intertidal Invertebrates of California* can be treated as something akin to a biographical dictionary. Knowing, or at least having a strong hunch about the name of the creature, the reader can here confirm that he has, indeed, probably seen this or that member of the community, and then learn some of what the experts know about it, as well as where to find further information.

A conspicuous contrast with a biographical dictionary is that no historian is ever likely, on the basis of some arbitrary legal point, to decide that henceforth Madame Curie will be known only as Marie Skłodowska. Many of my old friends in the intertidal zone, however, regardless of how well established their reputations were, are now assigned names different from those I learned 25 years ago. In some cases (e.g. *Mitella polymerus* and *Aletes squamigerus*) the older names are also included in the index to this volume; but for others (e.g. *Acmaea digitalis* and several congeners, and *Lepidopa myops*) there is no hint about former identity. This is, of

course, not so much a complaint about this book as about the problems inherent in nomenclature, and it can be argued that such changes represent "progress"; but I sometimes wonder. . . .

The neophyte will find the chapter introductions and the detailed text entries in this book most useful; the more advanced student will find the literature citations a good starting point for a research project; everyone will, I think, enormously appreciate the coloured plates. The professional biologist who knows some part of the literature on a particular group or species will no doubt note — as I did —

cases of mistaken conclusions, incomplete or outdated citations, and unevenness in coverage. But it would be petty to emphasize such deficiencies. The central point is that this book is an indispensable addition to the library of all who are interested in the common intertidal invertebrates of the California shoreline — and it contains a very attractive collection of colour portraits for those who will never get there. □

J.T. Enright is Professor of Behavioral Physiology at the Scripps Institution of Oceanography, La Jolla, California.

Old twists to an old tale

Jerry Donohue

The Double Helix: Text, Commentary, Reviews, Original Papers. Edited by Gunther S. Stent. Pp.298. ISBN hbk 0-297-77899-4; ISBN pbk 0-393-95075-1. (Weidenfeld & Nicolson/W.W. Norton: 1981.) Hbk £10; pbk \$5.95.

AN OLD proverb, as given by Gibbon, tells us that "happy is the nation that has no history". If this applies to substances, then DNA, surely, is the unhappiest of molecules, for it is the subject of innumerable biographies.

This new book contains: the complete *The Double Helix*; reprints of 12 reviews of it from 1968 and discussions of these by Gunther Stent; reprints of three papers from *Nature* by James Watson, Francis Crick and Aaron Klug; three letters to *Science* from 1968 by Max Perutz, M.H.F. Wilkins and Watson (see below); reprints of six papers from 1953–54 describing the original DNA work done in Cambridge and King's; and a name index the usefulness of which is somewhat marred by careless inaccuracies.

Should the fact that only about 5% of the contents of this new edition is previously unpublished material be cause for discontent? Probably not, as it brings together *The Double Helix* and some of the important writings it spawned. There are, however, other causes for discontent which are hardly contentious. The original text of *The Double Helix*, including the photographs and diagrams, ran to 233 pages; in this reprint these occupy 135 pages. This drastic compression was accomplished by abandoning the elegant format of starting each chapter on a new page; by taking the plates of the first (Signet, 1969) paperback edition, which was itself a compression of the original, and compressing it still further by putting more lines per page; and by reducing the sizes of the photographs, some by more than 50%, so that the individuals in the group photographs are virtually unrecognizable. Also reduced in size is the photographic reproduction of an historic

letter of Watson to Delbrück which is legible only with great difficulty and, in some places, totally undecipherable, a circumstance not caused by Watson's crabbed hand. Surely a book as important as this deserves a better production.

Something very mysterious should also be pointed out. In the diagram on p.111 the enol structure for thymine incorrectly has an oxygen atom where there should be a hydrogen atom. The same diagram on p.191 of the hardback first edition of *The Double Helix* is correct, and identical in every other respect, right down to the pattern of dots inside the rings of the bases. Now, admittedly, some of the early difficulties in cracking the DNA structure concerned the keto-enol business, but what sinister force, in Alexander Haig's terminology, took it upon itself to change a correct structure to an erroneous one?

There are several misconceptions and misreportings that ought to be dispelled before remarking on a few of the reviews included in this new volume.

● *The Myth of the Invention of Base Pairing.* Throughout the reviews runs the thread of the notion that Watson and Crick would never have discovered their structure had it not been for "rules" of Erwin Chargaff that in DNA (molar amounts of) cytosine = guanine and thymine = adenine. However, Chargaff himself had not yet referred to this relationship as rules, and, indeed, his own published analytical data gave merely the merest hint of these regularities, and he himself averred the possibility that the regularities might be accidental. Stent (pp. 168–169) has more to say on this.

The fault for these misconceptions must be squarely on the shoulders of Watson. We find him writing (p.75) "The moment [Spring 1952] was thus appropriate to think seriously about some curious regularities in DNA chemistry first observed by Chargaff", and then subsequently cheerfully disregarding them, and building a like-with-like, non-complementary, model. In addition, he had been assured by his friend, the biochemist Markham, that if

Cold Spring Harbor Laboratory have recently published the proceedings of the symposium on viral oncogenes which was held last year. The two-volume set of proceedings, comprising Vol. XLIV in the series *Symposia on Quantitative Biology*, is available from CSHL, Fulfillment Department, PO Box 100, Cold Spring Harbor, NY 11724, or through booksellers, price \$130 (\$156 outside the USA).

Review of James Watson's *The Double Helix* by J. Field

Hear the song of how the spiral
Complex, twisted, double, chiral
Was discovered. Its construction
Following some shrewd deduction
Was accomplished. From all sides "Oh
Tell it as it was" they cried so
Then he took his pen and paper
And composed the helix caper
Writing of the Cambridge popsy
Turned the world of science topsy
Turvy like had not before been
Done. The story is no more than
How the structure was discovered
Secret of the gene uncovered
Biological prediction
Better than a work of fiction
Stylistically breezy
Everything appeared made easy
And it was. The bases' pairing
Found at last — a small red herring
Notwithstanding. How he found them

And the hydrogens that bound them
Was a stroke more accidental
Than a work experimental
Chemistry was not his calling
He had read *one* book by Pauling
Models of the bases he'd made
With them on his desk top he'd played
Shuffling them and putting like with
Like he made no lucky strike with
Guanine, thymine, adenine. A
Letter came from Pasadena
Horrors! then the day was won for
Pauling's model was quite done for
Extra atoms he had set in
Where no atoms should be let in
With some phrases less than modest
Pauling's model's called the oddest
Back now to the basic pairing
How's the structure building faring?
Faster, faster goes the race, then
Everything falls into place when

Tautomers which nature chooses
Are the ones our author uses
Two chains (but here we can't be sure)
Plus this extra structure feature:
Pauling's outside bases inside
Pauling's inside phosphates outside
Now they all were quite ecstatic
Soon became they quite dogmatic
Having nature's secret later
Checked against the x-ray data
This was difficult, for Rosie
Found our scientist rather nosey
Bragg compares the book with Pepys. Is
This the verdict of Maurice? His
Thoughts alas we cannot gainsay
Trumpet blowing's not his forte
Tales like this do have a moral
Whether printed whether oral
Find the right man to advise you
Then you'll get a Nobel prize too.

Chargaff said that guanine equalled cytosine, he, Markham, was equally certain that it did not, because Chargaff's experimental methods inevitably underestimated the true amount of cytosine. On numerous other occasions Watson refers to Chargaff's "rules".

It is, on the other hand, abundantly clear from the narrative that these "rules" had nothing whatsoever to do with the construction of the DNA model. The model came first, and made the rules, not the other way around.

● *The Myth of the Undercover Agent.*

In some of the reviews selected by Stent the role which Peter Pauling played is displayed as not quite what it should have been. Thus Watson, according to Merton (p.215), reports "with temerity and self-mocking wit the occasions on which Linus Pauling's son, Peter, then a student at Cambridge, became a prime source of information about what his father was up to". Morrison (p.176) says that the idea for the double helix came "... out of hearing from the young American expert upstairs [actually, I was not upstairs at all, but in the same office as Bob Parrish, Watson, Crick and Peter] that the textbook was probably wrong, or out of rumors about what Pauling was thinking — relayed by his son Peter at Cambridge". Lear (p.195) goes considerably further in stating that Watson "discloses how he used his young friend Peter Pauling to spy on Pauling's brilliant father, Linus". And Lwoff (p.227) observes that "Peter Pauling works in the Cavendish, receives detailed letters from Pasadena and informs his colleagues of the evolution of his father's work".

On careful reading of *The Double Helix* we find that Watson says nothing of the sort. Consult, if you will, the name index for the citations to verify this, and come to the conclusion that these statements and innuendoes in the reviews are pure fiction. So much for spies!

● *The Myth of the Race.* Another

thread, nay, hawser, which twines through the reviews, is the "breathless race" between Watson and Crick on the one hand, and Linus Pauling on the other, towards that legendary ceremony, established by the dynamite tycoon, in the gloomy, sunless Stockholm of December. The same strand snakes its way through *The Double Helix*. How, though, can there be a race with only one team and no antagonist? At this stage (1952–1953), Pauling, whose reputation as a scientist was unequalled, had well over 200 publications under his belt, including a dozen or more books universally appreciated and widely read. His ticket to Sweden was long overdue. (He did, in fact, make this journey the very next year in 1954 to collect his first Nobel prize.) Was he racing Watson? Certainly not. To Pauling, in Peter's view (*New Scientist*, May 31 1973, p.558), DNA was just another chemical, intrinsically equal in interest to, say, sodium chloride; both presented interesting structural problems, so why race with DNA?

● *The Myth of the Filched Photographs and the Dipped Data.* The fourth and final myth concerns the X-ray data, and whether Watson and Crick came by them by unorthodox methods. Among the reviewers we find Sinsheimer referring to "cadged data" and a "privileged report", Lear to a "clandestine communication channel" and supposedly "confidential data", and Merton to "an odd expedient for gaining access to badly needed information". Lwoff and Chargaff (in his review, not included in this volume) quote a crucial paragraph from *The Double Helix* which led to both of them to comment on ethics. The same passage is reproduced by Perutz, the supposedly clandestine channel, in his letter, pp.207–210, in which, as well in the letters by Wilkins and Watson which follow, it is made abundantly clear that the report was not confidential, and the data had already been

disclosed in open seminars.

As a matter of fact, an X-ray diffraction photograph which, although not of the high quality of those produced by the two groups at King's, clearly shows the characteristic helical pattern of the B structure, had been published by Astbury years earlier, so all the fuss over stolen data was really unnecessary.

Five of the reviews of *The Double Helix* included here were previously reviewed by Gunther Stent, the editor. The article by Chargaff, that *vieux terrible* of the DNA saga, although not reprinted because he withheld permission, is reviewed. (Anyone wishing to read the original will find it in *Science* 29 March 1968; p.1448). Of the seven new reviews, one of them warrants comment, that by the bipseudonymenous "F.X.S.", originally "F.R.S.". This is not a review at all, but verbal virtuosity by a polylingual sesquipedalianist, and why Stent chose to include it is an enigma, as it adds little insight into DNA or those connected with it. F.X.S. is coy about his identity, but really, who cares? His piece is worth reading, but hardly reprinting.

There are numerous other important reviews not included by Stent — by Lord Todd in *Chemistry in Britain*, Sokolov in *Newsweek*, Bernstein in *The New Yorker*, Maddox in *Nature* and so forth. Another not selected by Stent, by J. Field, appeared in the *Journal of Irreproducible Results*, (Vol.17, December 1968, p.53). Because it expresses as well as anything the flavour of Watson's book, it is reproduced above (with the permission of the editor).

Anyone who hasn't yet read *The Double Helix* should certainly do so, but, even if you own a copy, the additional material in the present volume amply justifies its purchase. □

Jerry Donohue is Rhodes-Thompson Professor of Chemistry at the University of Pennsylvania. He spent the academic year 1952–1953 as a Guggenheim Fellow at the Cavendish Laboratory, University of Cambridge.

Schumacher reconsidered

Eric Ashby

Small is Possible. By George McRobie. Pp.331. ISBN hbk 0-224-01858-2; ISBN pbk 0-06-090694-4. (Jonathan Cape/Harper & Row: 1981.) Hbk £7.95; pbk \$5.95.

THERE is a mysterious jackpot logic about best-selling books on the environment. Why has Vogt's *Road to Survival* sunk without trace while Leopold's *Sand County Almanac* is quoted and requoted? Both were published a generation ago; both are elegantly written. Why is Thoreau, who liked his wilderness to be as tame as a tabby cat, still regarded as a conservationist's guru? And why is Schumacher's *Small is Beautiful* treated as a seminal tract in self-sufficiency? Surely Tolstoy had similar ideas and there were Tolstoyan communities (well described by W.H.G. Armytage in *Heavens Below*, published 20 years ago) in Britain in the 1890s.

Small is Beautiful, I'm told, still sells well, thanks (ironically) to the large-scale techniques of book production and marketing. It is seductively written and full of wisdom. But wisdom, alas! is only a minor ingredient in practical politics. As devotional reading, Schumacher's book is still popular; but as a blueprint for a new economic policy in the manifesto of a political party (except, perhaps, the misnamed ecology party) it's not a starter. Not only in the capitalist world but in the communist world too, small may well be beautiful, but it is the giant enterprise that provides energy, packaged foods in the supermarket, communications, transport, the dish-washer in the kitchen and the car in the garage. The communes in China are an exception to this generalization, but will they survive as China becomes a great industrial power? For Britain, at any rate, it does not seem unduly cynical to assume that if a choice had to be made at the polls between an economic life style which was, in Schumacher's sense, beautiful, and one which gave the greatest material comfort to the greatest number of people, then beauty would lose out.

So the stock comment on *Small is Beautiful* is "beautiful, no doubt; practicable, no". Therefore when a sequel appears with the title *Small is Possible*, it should be accepted as a challenge to re-examine Schumacher's thesis. It is in that spirit that I read the book.

George McRobie, the author, was Schumacher's assistant and co-founder of the Intermediate Technology Development Group, ITDG (one of the many — far too many — acronyms in the book). He does not pretend to write with the felicity of Schumacher but he doggedly assembles evidence to validate the book's title. The book is, for the most part, an annotated catalogue of enterprises in intermediate or

"appropriate" technology which are consistent with Schumacher's thesis, namely that conventional capital-intensive projects are a disaster in Third World countries and may well be the signpost toward disaster in industrialized countries; and — this is the message of the book — that there are alternative small people-intensive projects which are viable. McRobie gives a list, with addresses, of nearly 100 organizations concerned with appropriate technology engaged in hundreds of little experiments. The evidence persuades me (at any rate) that there is a minority movement to reject the life-style of what used to be called the military-industrial-complex; not on political grounds but on what may become a much firmer foundation: moral grounds. It is no more than a sprinkling of enthusiasts (some of them cranks) in the interstices between super-highways, power stations, acres of petrochemical works and refineries; uncoordinated, fragmented. But something interesting is going on, and if the fragments coalesce a new life style might emerge.



The late E.F. Schumacher, prophet of self-sufficiency.

What is going on? Here are a few samples taken from McRobie's book.

He deals first with the most urgent and relevant application of Schumacher's thesis: the design of equipment for agriculture and industry in Third World countries. The principle is simple. In countries with surplus manpower, and without technical sophistication, capital investment per worker has to be as low as possible. When a British trade mission went to Nigeria, Schumacher's group (ITDG) put into their hands a stencilled *Guide to Small-scale Equipment for Rural Development*, "of animal and man-operated agricultural equipment obtainable from little hole-and-corner firms in Britain". Not at all what the mission intended to propagate in Nigeria; but (said Schumacher) "This thing was torn out of

their hands by people who said: For the first time you bring us something of interest, instead of the usual glossy catalogues". That was 15 years ago. Since then the ITDG has had an impressive record of service to Third World countries. A building panel set up, in five African countries, units for making bricks at a capital cost of £400 per workplace, compared with £40,000 per workplace in a large modern brickworks; and it invented and put into production a fibre-reinforced cement for roofing at a fraction of the cost of the normal commercial product. A water group published handbooks on the making of pumps, wells and low-cost irrigation systems. A farming group has designed and put into production (in Zambia and Nigeria) workshops for making machines for weed-control, groundnut harvesting and fibre stripping. A transport unit based on the Engineering Department at Oxford has designed the Oxtrike: a cycle rickshaw which can be made in Third World countries. Windmills in Ethiopia, water pumps in Nepal, mini-hydro generators — all locally made — replace the imported equipment which cannot be kept in good repair without skilled technicians. And — most imaginative of all — the Rural Health Panel has, under the skilled guidance of Dr Katherine Elliott, promoted the training of auxiliaries in health care, in collaboration with the World Health Organization.

These efforts may seem to be trivial crumbs to offer the Third World when set beside the massive expenditures on steel mills for India or on the Volta Dam for Ghana; too trivial for all the effort involved. But they have an importance which cannot be quantified: they may, with luck, convince the Third World that it does not want to ape Europe and America in its patterns of society. And it may help to repair some of the damage we have done in exporting the wrong kinds of goods and education to the Third World. Thus, when we exported medical education to the university colleges of tropical Africa we arranged (with the noblest of motives) that its quality and content would be indistinguishable from that in a London medical school. The surgical staff, training doctors skilled to remove the gallstones of cabinet ministers in the capital city, regarded with impatience the activities of an African professor of social medicine, who put his efforts into the holding of competitions in the villages for the construction of earth closets. Anyone who dared to suggest that the prime need was for medical auxiliaries (as the Soviet Union and China had) was scarified with criticism. The universities were suspicious of any subject in the curriculum which would not have been appropriate in London. One encouraging message in McRobie's book is that this technological snobbery is diminishing. African universities are becoming relevant to African problems. Countries where Small has been

Practicable (if not Beautiful) for millennia may still have a chance to escape some of the social perils of modern technology.

What about the industrial West, where the proudest boast of technologists is that they save manpower at the cost of capital investment per workplace? It would be stupid to dismiss the enormous social benefits of mass production. For every one man whose life is dissipated on a production line in British Leyland there must be hundreds of families whose lives have been liberated by the possession of a car. Holidays, picnics, visits to friends and relations, life in a suburb and a short drive to the office: all these are incomparably easier to secure than they were for the so-called lower classes when I was a boy. While there was full employment, and material progress brought with it some improvement in the quality of life, the achievements of the industrial West were welcomed. For the first generation owner of a car and a refrigerator and an air ticket to Majorca, the warnings of Schumacher were dismissed as neurotic.

It is a different story today. Thanks to education and the media, the unemployed are more articulate than they were in the 1930s. They do not equate unemployment with leisure, even though they are paid to be unemployed. To be without work is as dangerous as to be without love. So, suddenly, the application of Schumacher's ideas (and not just Schumacher's: they are the ideas of William Morris and Tolstoy in modern dress) becomes relevant and politically important. McRobie describes how hundreds of small groups of people in America, Canada and (on a smaller scale) in Britain have detached themselves from the conventional pattern of society, to seek a pattern which does not have the formula "economies of scale" woven into the design. In America some of the experiments have official support. There is a National Center for Appropriate Technology, inspired by Senator Mike Mansfield. There is the California Office of Appropriate Technology, started by order from Governor Jerry Brown, with a budget of \$900,000, encouraging the use of solar energy, and designing small-scale sewage systems and techniques for aquaculture. In the Bronx, some of which is a waste of derelict city lots, an Institute for Local Self Reliance has mobilized local labour to cultivate gardens and rehabilitate decayed property. And in Canada — in many ways the most enlightened of Western nations — there is in the ugly mining town of Sudbury a project called "Sudbury 2001" which, with a provincial government grant of \$600,000, is preparing some of the workforce for the inevitable redundancies they see ahead.

McRobie's chapter on the "alternatives" movement in Britain is less encouraging. It seems (perhaps there is another side to the story) that neither governments nor trade unions have taken much interest in the possibilities of small

collectives. The experiences of the Kirkby Manufacturing and Engineering Ltd (KME) and the *Scottish Daily News* do not generate much confidence. Of course it is sheer fantasy to imagine that ICI, BP and the Central Electricity Generating Board will dissolve in a flood of little collectives making their own paints, heating their farms with home-made methane, using windmills knocked together by the local blacksmith. But the unemployed displaced

by a combination of world diseconomy and new technology are going to demand work. If they do not get work, society as we know it may be shattered. If we have to invent work not to produce things for profit but to satisfy a deep genetically coded human need, then perhaps Schumacher's ideas are worth another thought. □

Lord Ashby is Chancellor of Queen's University, Belfast, and a Fellow of Clare College, Cambridge.

Science fiction as Marxism

James Gorman

Robert A. Heinlein: America as Science Fiction. By H. Bruce Franklin. Pp.232. ISBN hbk 0-1950-2746-9; ISBN pbk 0-1950-2747-7. (Oxford University Press: 1981.) Hbk \$18.95, £10; pbk \$4.95, £3.95.

READING H. Bruce Franklin on Robert Heinlein is a bit like reading an analysis of a sexual farce by a stiff-necked, bible-bound fundamentalist who can never get his mind off the fact that sex is, after all, sinful. Franklin has produced a painstakingly detailed Marxist analysis of what, at its best, is flashy escapist fiction. Heinlein's heroes, his cynicism and sentimentalism, his taste for adventure, all his tales of space exploration, war, sexual fantasy, time travel and immortality become a rather dreary reflection of the contradictions of capitalism in the United States.

Franklin leaves the science in the fiction more or less alone; this omission does no disservice to Heinlein. Science only provides him with the authority by which reality is breached — it is the excuse for the reader to believe in the fantasy. When it suits his purpose, Heinlein is not averse to using magic in its place.

Heinlein is at least as fond of social theory as he is of science, and his books are full of expostulation on one half-baked political philosophy or another. And, as Franklin points out, he does reflect his times. In the 1950s he produced anti-communist novels; in the 1970s he explored narcissistic sexual excess in some of the worst writing of his career. And, in 1961, at the beginning of one of America's most notorious decades, he produced what is probably his most popular book, *Stranger in a Strange Land*. In it a young human being, raised on Mars, by Martians, comes to Earth, views us all through alien, though human eyes, and finally founds a religion that is part universal sexual love, part hucksterism, part anarchism and part violent elitism. The gentle, loving hero is in the habit of making his enemies, and other undesirables, disappear by a wink of the mind, a trick he learned on Mars.

This is, indeed, very close to "America as Science Fiction". But when Franklin presses his point to argue that the hero "embodies our most infantile fantasies and

the central goal of bourgeois ideology — the unfettered freedom of the individual will", he is less convincing. Indeed, when, throughout the book, he holds up Heinlein as a mirror that reflects the nature of capitalist society, he undermines his case by two unfortunate omissions — he leaves out Heinlein, and he makes his analysis on the basis of unargued and undemonstrated assumptions.

Where indeed is Heinlein in all of this? Franklin gives us only a thumbnail biography, which concentrates more on the economic conditions in the country when the author grew up than it does on Heinlein himself. And where is Heinlein's imagination, his writing skill, his ability to create compelling stories? A writer of fiction is less a mirror than a lens, that refracts rather than reflects the light of his culture, bending it with his own idiosyncrasies. In Franklin's formula the curve of Heinlein's lens is missing.

Also missing is some back-up for his ready-made Marxism. It may well be true, as he notes in passing, that the conditions of the two main groups of workers in America before World War II were "the debt peonage characteristic of the majority of Black rural workers and the wage slavery of mine and factory labor", and that "class struggle . . . has constituted actual human history", but his case would be stronger if he demonstrated his points rather than asserting them. Nor are these isolated examples. He tells us, for instance, that "If the will is free to do anything it wishes, the will is free from the apparent laws of the physical universe and also free from the apparent laws of human social development". He doesn't tell us what either set of laws are, perhaps assuming that the precepts of dialectical materialism are as well known and widely accepted as gravity.

What Franklin's book could use, finally, is a bit more of an internal dialectic. If we were to see the clash of culture and individual in Heinlein's work, we might come away with a richer understanding of the author, and of America. □

James Gorman is a staff writer on Discover magazine.

Spring Books

Gel Electrophoresis of Proteins: a practical approach

Edited by D. Rickwood and B. Hames

A laboratory aid for all those wishing to separate proteins by gel electrophoresis. Contains detailed recipes and step-by-step instructions on the procedures involved in all the different types of electrophoretic separation.

0 904 147 22 3 (soft) £7.50/US \$18.00
300 pp (approx.)

Rapid and Automated Methods in Microbiology and Immunology: a bibliography 1976-1980

A unique reference tool to help you find the simplest method or technique best-suited for your laboratory. Contains over 3000 citations on practically all the rapid and automated techniques introduced in the field during the last five years.

0 904 147 19 3 (soft) £15.00/US \$36.00
200 pp (approx.)

Tooth Surface Interactions and Preventive Dentistry

Edited by G. Rolla, T. Sonju and G. Embery

Proceedings of a workshop held December 1980 in Oslo and part of a series devoted to current aspects of dental research concerned with the interactions occurring at the tooth surface. Particular emphasis is directed at the role of inorganic ions, especially stannous fluoride, in influencing the metabolism of oral bacteria and their adherence properties.

0 904 147 29 0 (soft)
£15.00/US \$36.00 217 pp, 40 illus.

Immunology of the Eye Proceedings of three workshops sponsored by the National Eye Institute, USA

Workshop I: Immunogenetics and Transplantation Immunity
0 904 147 25 8 (soft)
£10.00/US \$25.00 300 pp

Workshop II: Autoimmune Phenomena and Ocular Disorders
0 917 000 07 2 (soft)
£10.00/US \$25.00 200 pp (approx.)

Workshop III: Infection, Inflammation and Allergy
0 917 000 08 0 (soft)
£10.00/US \$25.00 300 pp (approx.)

Prepaid orders and requests for further details to:

Information Retrieval

1 Abbey Street, Eynsham,
Oxford OX8 1JJ, England
1911 Jefferson Davis Highway,
Arlington, VA 22202, USA

Circle No.31 on Reader Enquiry Card.

Colourful tale of two Voyagers

J. Veverka

Voyage to Jupiter. By David Morrison and Jane Samz. Pp.199. (National Aeronautics and Space Administration: 1980.) Pbk \$7.50 US only.

NASA-sponsored books describing flight missions to various planets are not hard to find. Unfortunately, good ones are as scarce as impact craters on the surface of Io. Typically, such books amount to little more than collections of pictures intermingled with spacecraft details and other mission minutiae. Science gets mentioned only in passing and at so superficial a level that even the casual reader feels short-changed.

Voyage to Jupiter is not such a book. It is a beautifully illustrated, vivid account of the two historic Voyager encounters with Jupiter in 1979, as well as a factual summary of the initial scientific results. Everyone should find it interesting; even the sophisticated reader will find it informative.

In part, the successful outcome may result from the combination of talents of the two authors: Samz, a science writer, and Morrison, a planetary scientist and a member of the Voyager Imaging Team who was personally involved in many of the discoveries described in this book. Certainly, this volume does contain details of the spacecraft construction, the mission profile and the requisite pictures of the major participants, but interwoven with all this is real, substantial science. Voyager's many discoveries — rings, volcanoes, lightning, new satellites and so on — are not merely catalogued chronologically, but are tied into a comprehensive new picture of Jupiter and its satellites. Without question, *Voyage to Jupiter* is more than a NASA picture book, but it is certainly that too. The Voyagers are the first planetary spacecraft to have sent back large quantities of true colour data, and it would be difficult to imagine a better subject for colour imaging than the jovian system. Many of the best of these pictures are reproduced admirably in this volume.

The book opens with a pair of chapters summarizing our pre-Voyager knowledge of the Jupiter system based on three centuries of telescopic observations and the initial reconnaissance by Pioneers 10 and 11 in 1973-1974. Next, the authors discuss the spacecraft, the scientific work that its eleven instruments were designed to carry out and some of the individuals responsible for these experiments. Then come three chapters giving a day-by-day account of the two encounters. Here the reader can share some of the excitement experienced by those who were fortunate enough to be in Pasadena during these two historic events when major discoveries were being made almost hourly. Two of the remaining three chapters aptly summarize our

current, post-Voyager, view of the planet and its satellites, and the volume ends with a discussion of future missions to Jupiter. Sadly, this chapter must be very short. At present, the only mission planned is NASA's Project Galileo, which will investigate the jovian system in the latter half of this decade. For a while at least, our view of Jupiter and its satellites will remain that relayed to us by Voyager in 1979.

Voyage to Jupiter — an excellent little book — sets a high standard for future NASA publications of this sort. In these days of \$30-\$40 science books, it is an absolute bargain which I heartily recommend. □

J. Veverka is at the Laboratory for Planetary Studies, Cornell University, New York.

Here comes Halley

David W. Hughes

The Comet is Coming! The Feverish Legacy of Mr Halley. By Nigel Calder. Pp.160. ISBN 0-563-17859-0. (BBC: 1980.) £8.75. To be published in the USA by Viking, \$12.95.

"EVERY seventy-six years or so Halley's comet punctuates history like an exclamation mark in the sky". It was last seen in 1910 and it is due back soon. To be precise it will be at perihelion, the point in its orbit closest to the Sun, on 9 February 1986. The preparation for this momentous event has already started. The European Space Agency is busily organizing the Giotto spacecraft which will pass as close as possible to the cometary nucleus. The International Halley Watch is being supervised by the Jet Propulsion Laboratory in Pasadena and this will co-ordinate all the ground-based and near-space observations of the comet. The public are getting ready

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Edmund Halley, Astronomer Royal from 1720 to 1742.

too. Brian Marsden once said "To the man in the street the Solar System consists of Mars, the rings of Saturn and Halley's comet". Comets fascinate people and the next few years will see a flood of books, television programmes and newspaper articles on the subject.

Nigel Calder and the BBC are in the vanguard and *The Comet is Coming!* is the book of the television programme that is to be shown first on BBC2 in mid-1981. The book is excellent. Calder has a beautiful turn of phrase and often had me chuckling. The text is non-mathematical and infectious readability. The illustrations have been carefully selected and many are new to me — not for Calder is the regurgitation of the old faithful. The style is journalistic but you are never left in doubt that the author is a sensible scientist.

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Portent of a new King? Detail from the Bayeux tapestry showing the appearance of Halley's comet.

Halley's comet and its like are ostentatious and elusive and because of this they create a no-man's-land of speculation where crazy ideas may be fantasies or brilliant insights. This book is a sure guide to this no-man's-land and underlines the joys, mysteries and frustrations of cometary science. We read of the attitude of the superstitious who regard comets as telegrams from the Gods — disorders in heaven importing changes of times and states. The role of Edmund Halley is stressed, especially his work on orbits. The physical and chemical nature of comets and the suggestions as to their origin are reviewed, as are the problems of cometary discovery and observation. Fred Hoyle and Chandra Wickramasinghe's theory that comets could have brought life to Earth and are even today scattering disease over the globe through the influx of dust is discussed. Calder examines the consequences of a collision between the Earth and a comet; and special reference is obviously given to the demise of the dinosaurs about 65 million years ago!

The book is well referenced, well indexed and I recommend it strongly. □

David W. Hughes is a Lecturer in Physics and Astronomy at the University of Sheffield.

Star-spangled dream

Brian O'Leary

Bound For the Stars. By Saul J. Adelman and Benjamin Adelman. Pp.335. ISBN hbk 0-13-080390-1; ISBN pbk 0-13-080382-0. (Prentice-Hall: 1981.) Hbk \$17.95, £11.65; pbk \$8.95, £5.80.

"WE ARE too great a nation to limit ourselves to small dreams" said President Reagan in his inaugural address. Amongst the budget slashing that has followed, it is difficult to imagine just what dreams Mr Reagan has in mind.

One thing is clear: there is no shortage of big dreams in space. Sound engineering concepts wait for their political and commercial opportunity. Once they are triggered, we can expect to witness an explosive renaissance in the high frontier. The question is not whether it will or should happen; it is when, how and by whom. Soviet-American political and military pressures mean we are in space to stay.

Different writers, however, have different goals. Gerard K. O'Neill, in *The High Frontier*, argues the case for satellite solar power stations and space habitats manufactured in space from lunar materials. Carl Sagan, in *Cosmos*, describes the excitement in expanding scientific programmes in modern astronomy, the search for extraterrestrial intelligence and unmanned planetary exploration. In *The Fertile Stars*, I show how asteroid mining and retrieval could provide the key to expanding the limits to growth of the world supply of energy, food and raw materials, to pave the way to space colonies and manned planetary exploration. Saul and Benjamin Adelman have a rather less accessible dream to offer; *Bound for the Stars* champions interstellar travel as the primary goal for future space activity.

There is little question that NASA's lack of a well-defined objective has hurt the American space program immensely. The lunar-landing commitment made by President Kennedy in 1961 (and completed eight short years later) was the last grand goal of space. We are now seeing the final vestiges of Apollo momentum petering out — the Voyager Saturn fly-bys will be the last American planetary probes to return new data for at least another five years.

Apart from the space shuttle, the US civilian space programme is virtually dead, while the Soviets, Europeans and Japanese move steadily forward. For example, all three plan a Halley's comet mission during its once-in-a-lifetime appearance in 1985 and 1986; budget cuts have grounded an American Halley probe. The Adelmans say these retrenchments can be ascribed to shortsighted year-to-year planning in the US federal government, and contrast this with the project-approval system of the European Space Agency. They stress the importance of changing NASA's charter to protect it from the lack of continuity it

suffers from the annual budget cycle.

The best feature of the book is its wealth of detail on concepts of interplanetary and interstellar travel. The authors trace the history of the development of propulsion systems and navigational techniques with skill and perceptiveness. Most of the concepts have not flown. They describe, for example, the ill-fated Orion Project (in which exploding atomic bombs propel a large spaceship) as a "science orphan no one wanted to adopt".

The Adelmans' stand that interstellar travel should be the basic unifying aim for future space activity is controversial. Because starships are well beyond our current technological capability, many decision-makers will not accept such a heady goal. By contrast, Apollo was — and manned Mars landings, asteroid mining, space industrialization, space colonies and satellite solar power are — feasible projects which could use today's technology.

The first few chapters of the book are a smorgasbord of what appear to be afterthoughts. The authors skip rapidly from the origin of life to the search for extraterrestrial intelligence to a survey of the Solar System, back to life in space, plans for biological experiments aboard Spacelab, controlled ecologies, Solar System travel, space colonies and so forth.

These sections are weak, covering the subjects incoherently with misleading and biased remarks: for example, the belief that all life must be carbon-based and depend on water (see *Life Beyond Earth* by Gerald Feinberg and Robert Shapiro for a different view); "Dr Carl Sagan of Cornell University believes that about a million civilizations exist in the Galaxy on the same or more advanced level than Earth civilization" (p.17, too strong); "The galactic cosmic rays from outside the solar system are not a serious hazard, even for voyages lasting more than a year" (p.41, not true). In addition, they show a strong bias against satellite solar power and for nuclear power (pp.80-85) which, in the United States, might prove politically unacceptable.

The real strength of the book, then, is in its thorough discussion of the prospects of interstellar travel. I have been looking for a handy reference on the subject, and this is most certainly the best available. I applaud the Adelmans' remarks about the importance of keeping the space programme alive and that we need a goal to work toward.

I believe political and military pressures are the foot in the door, economics will become the driver (space industrialization) and the thorough exploration of the Universe will be the ultimate and most fulfilling by-product.

In spite of all the budget-cutting and pessimism, we have some ideas. The Adelmans' treatment of interstellar travel in this new book is one of them. □

Brian O'Leary is a former scientist-astronaut, and author of *The Fertile Stars*, to be published in June.

The pluviiculturists

Richard W. Katz

The Rainmakers: American "Pluvi-culture" to World War II. By Clark C. Spence. Pp.185 ISBN 0-8032-4117-8. (University of Nebraska Press: 1980.) \$15.95, £9.60.

In 1930, "Dr" George Ambrosius Immanuel Morrison Sykes was employed by the Westchester Racing Association to ensure sunny weather at Belmont Park racetrack in New York City. Although not a meteorologist, Sykes called himself a "meteorologist", that is, someone who can actually control the weather. Besides suppressing rainfall, Sykes' Weather Control Bureau also claimed to be able to prevent snow, hail, frost and hot weather. The Bureau could, on the other hand, generate rainfall if so desired. To perform these seemingly impossible feats, Sykes relied upon a bizarre assortment of equipment, ranging from a toy propeller to a tub of smelly chemicals. His efforts to guarantee sunshine at Belmont Park, needless to say, met with only mixed success.

Although the putative beginning of scientific weather modification dates from 1946, with Vincent Schaefer's successful seeding of clouds using dry ice, Clark Spence's entertaining book describes the largely forgotten history of earlier, less-scientific rainmaking attempts in America. David Jordan coined the term "pluvi-culture" to refer to the endless rainmaking schemes that were inevitably proposed upon the occurrence of drought. Most of these "pluviiculturists" were not nearly as outlandish as "Dr" Sykes. They included James Espy, a true scientist, whose theories concerning a relationship between fire and rainfall were respected by the scientific community. The most popular rainmaker was probably Charles Hatfield (the "miracle man"), a former sewing machine salesman, who used the "smell maker" technique involving the release of chemicals into the atmosphere. Reputedly responsible for the great flood of 1916 in San Diego ("Hatfield's flood"), he has become a part of the folklore of southern California.

The recipe for success as a rainmaker was simple. Above all, it required being a convincing salesman to take advantage of the public's gullibility. When faced with drought, farmers were desperate and

Two psychology textbooks recently reviewed in *Nature* (289, 710; 1981), *Behavioral Neuroscience* (by C. W. Cotman and J. L. McGaugh) and *Physiological Psychology* (by T. S. Brown and P. M. Wallace), are available in the UK in a cheaper international edition. Both are published by Academic Press, and the prices of the international editions are £6.25 and £6.40, respectively.

willing to try almost anything. This "drought psychology" made the "no rain, no pay" schemes offered by the rain-makers highly attractive. Aided by weather forecasts and information on local meteorological patterns, the pluviiculturists depended on occasional coincidences with the occurrence of natural precipitation to demonstrate the viability of their rainmaking operations. Plausible excuses could always be dreamed up in case of failure. Sensationalized journalism helped too, publicizing apparent successes, while not bothering to mention failures. The scientific community, including the US Weather Bureau, remained steadfast in opposition to nearly every rainmaking scheme, but to no avail; scientific expertise was no match for expert salesmanship.

The rainmakers took advantage of several common fallacies, all somewhat statistical in nature. The *post hoc, ergo propter hoc* fallacy arose in connection with the coincidence between rainmaking operations and subsequent rainfall. Pluviiculturists always took credit for any rainfall, even if it occurred several days after their operations or only at distant locations. Another fallacy concerned the often misused "law of averages". Rainmakers were well aware of the odds for rain, correctly realizing that it was sure

to occur eventually. That the public was confused about the law of the averages is no wonder; even Spence is apparently confused about it. He suggests that the task of a rainmaker was made easier, because an extended drought increased the likelihood for rain.

It is impossible to read Spence's book without being compelled to make comparisons with more recent weather modification activities. Some striking similarities are present. Overstatement continues to be employed to sell rain-making, with one current euphemism for such activities being "weather resources management". Journalists' reporting of the subject remains somewhat misleading or inaccurate, and the public tends to accept the same fallacious reasoning. As always, interest in pluviiculture is revived whenever a drought occurs. To justify government funding of weather modification programmes, politicians rely to some extent on the same "it's worth a try" or "it can't hurt" arguments. If Spence's history of American rainmaking were extended to the present, how different would it be? □

Richard W. Katz is in the Department of Atmospheric Sciences at Oregon State University, Corvallis.

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

CHAPPIE. "This is great. Maud wants me to take her to the foot-ball game. I'll just buy a thundershower and knock out the whole business."

Rainmaking Humour, from *Harper's Weekly* October 24, 1891.

Richard of York's battles not in vain . . .

Ann Henderson-Sellers

Rainbows, Halos, and Glories. By Robert Greenler. Pp.240. ISBN 0-521-23605-3. (Cambridge University Press: 1981.) £15, \$24.95.

ACCURATE recollection of the ordered colours of the rainbow by means of the nursery sentence, "Richard Of York Gave Battle In Vain", is a necessary but insufficient basis for understanding in the eyes of Robert Greenler. He wants us also to be moved to wonder at the mysteries of optical phenomena in the sky. His book is reminiscent of a fascinating, extended tutorial with a professor of physics who, fortunately, combines hard scientific research with an almost childlike delight in beautiful things. During his discourse we are prompted to exclaim at magnificence, explain apparent enigmas and reinterpret historical events through computer simulations of atmospheric optics.

The book divides into three sections dealing with rainbows and ice-crystal optics, and then simulation and understanding of historical and modern sky displays, before finally returning to colourful mysteries, myths and mirages. Throughout, facts and frustrations are, like the individual rainbow we each view, enriched by Greenler's personal experiences. For instance, he illustrates the concept of the constant angular size of rainbows by a description of his doomed attempts to photograph the, apparently, miniature bow formed in the spray of his garden hose. The author is clearly "hooked" on his subject; he admits to seeing and searching for rainbows everywhere and has even captured infrared bows on film.

Despite his obvious love of rainbows it is with the optical displays formed by ice-crystals that Greenler is most strongly associated. As a physicist who has spent extended periods in Antarctica and the Arctic his fascination with this topic is understandable, but it is frustrating to realize that many of the sky effects he illustrates in this section are seen only rarely in more populated regions. This is because the huge variety of optical effects discussed depend upon the presence of simple hexagonal ice-crystals while most of us see our ice-crystals as snowflakes. This pinpoints the only real criticism that I can make of the book: it is not focused on a specific readership. As a meteorologist I am frustrated by the fact that the weather which, after all, is creating most of these effects, is mostly ignored. This is a pity because it makes an otherwise outstanding book tricky to use directly in taught courses. Despite the fact that the book is not intended as a textbook, with little additional effort topics could have been related to other disciplines. For instance, the very thin cirrostratus clouds

responsible for the 22° halo, itself generally considered as a forerunner of stormy weather, could also be considered to be a sky phenomenon. I can testify to having used halo occurrence to convince a non-meteorologist that a blue sky could be totally cloud-covered. Despite these reservations I was drawn on by the enthusiasm of the writing. Indeed, I spent some time testing the author's assertion that hexagonal ice-crystals have preferred orientations in the atmosphere, to the extent of spinning and dropping computer cards around my office.

With a flair for detective work which Holmes would have envied, Greenler considers his computer simulation of the famous "Arcs of Lowitz". These composed part of the historic sky display seen and sketched by Tobias Lowitz in St Petersburg in 1790. The accuracy of the simulated curves seems remarkable but Greenler remained dissatisfied. By returning to the written description he has demonstrated that the sketch is not a "snapshot" but a composition of a number of hours of viewing. This revelation solves the outstanding problems

which depended upon knowledge of the solar zenith angle.

The final 50 pages are devoted to colourful optical phenomena in both senses of the word. We are treated to pictures, and explanations, of white and black clouds in the same scene and then yellow, green, blue and red clouds. Such comparatively "every-day" sights pale before the "Specter of the Brocken". This apparition takes the form of a misty giant around whose head are a series of brilliant halos. Many air travellers have seen the same effect, a glory, which simply requires the observer to be in direct sunlight with his anti-solar point projected onto a cloud.

In the opening pages of the book the author states, "I have chosen, knowingly to include myself . . . my delight at seeing . . . my excitement in the process of discovery". Despite my slight unease at such an apparently selfish exercise I shall use and recommend the text. Like the Pied Piper of Hamelin in his rainbow-coloured coat, Greenler charmed me into following and will, I am sure, similarly entrap other readers. It appears that the white rose of the house of York was not lost in vain. □

Ann Henderson-Sellers is a Lecturer in Meteorology at the University of Liverpool, currently studying global climate at the Goddard Institute for Space Studies, New York.

Evidence of intent

John Stachel

The Greatest Power on Earth: The Story of Nuclear Fission. By Ronald W. Clark. Pp.342. ISBN 0-283-98715-4. (Sidgwick & Jackson: 1980.) £8.95. To be published in the USA by Harper & Row, \$13.95.

SHORTLY after the accident at Three Mile Island, a Japanese journalist visited a centre for evacuees from the vicinity of the reactor. As startling as the accident itself was the fact that a reference to Hiroshima drew a blank from those he interviewed; none of them recognized the name. Such ignorance in itself justifies more books on the story of the atomic bomb, which is basically the topic of Clark's book.

Over two-thirds is devoted to the story, from the discovery of atomic fission in late 1938, through the wartime efforts to develop nuclear weapons in Britain, the United States, Germany and Russia, up to the explosion of the first Russian bomb in 1949. A 40-page prologue to the book attempts a potted history of nuclear physics; while the last 40 pages wind up the story with brief accounts of the development of the fusion or hydrogen bomb, the attempts to use fission as an energy source and the growing awareness of radiation perils associated with peaceful and military uses of nuclear power.

The prologue is not too successful. It betrays a shaky acquaintance with the

scientific literature, resulting in numerous small errors and confusions. The placement of emphasis, too, is sometimes curious.

The core of the book demonstrates a wide acquaintance with the atomic bomb literature, both original documents (including some discussed in print for the first time) and later memoirs and analyses. On the whole, it seems a thorough and reliable account, correctly giving greater emphasis to work done in Britain than many American sources. This portion of the book commends itself to those in search of a readable, fairly non-technical account of the scientific, diplomatic and military events leading up to the "Birth of the Bomb" (to use the title of Clark's previous account of Britain's underrated contribution to the early phases of the story). Those in search of a deeper analysis of the potential background to the decision to drop the two bombs on Japan are still advised to consult Martin Sherwin's *A World Destroyed* (Knopf, 1975).

The frequent use of hyperbole is annoying: "In 1905 Einstein erupted into a surprised world" (elsewhere "an astounded scientific world" receives Einstein's 1905 relativity paper); Russian nuclear research is interrupted in 1941 because "German tanks with muffled treads moved through the darkness across the Russian frontier", and so on. There are also a few curious phrases: "The predominance of Jewish nuclear physicists on the Continent — in strong contrast to the situation in Britain —

has long been a subject for discussion"; by whom one wonders?

But the more serious problems of the book begin with its title. It is basically the story of nuclear fission, as its subtitle indicates, yet the title clearly refers to nuclear fusion — a reference reinforced by the exploding hydrogen bomb on the jacket cover. Careful reading of the text makes clear the distinction between fission and fusion, but hasty readers may be left confused.

More important is the question of the distribution of emphasis inherent in the structure of the book. Surely, the critical element in our current nuclear predicament is the combination of the fusion or hydrogen bomb, with its capacity for unlimited growth in explosive power, and a missile technology capable of accurately placing such bombs on target within minutes of launching. But no sense of the crucial nature of this development emerges from the brief concluding section of the book. Indeed, what is the moral of Clark's story? He says all the "right" things about the perils of nuclear weapons, and dangers of nuclear radiation — except for one sentence in his concluding paragraph, which plays into the hands of those currently trying to prepare the public for "limited" nuclear wars. He says,

It would be tragic if a limited nuclear exchange, halted after each side had experienced the agony of one city put to the bomb, were necessary to bring men to their senses and to realize, at last, that there are no victors in nuclear war.

Not only would it be tragic (reminding one of the man being led to his execution who assures the audience that this will certainly teach him a lesson!); it is so unrealistic a prospect, as Lord Mountbatten came to realize, that its contemplation as a pedagogical device seems irresponsible.

Finally, there is the broader question raised by the context in which Clark's account places the story of nuclear weapons. Focusing attention in the opening sections exclusively on the development of nuclear physics fosters the all too prevalent view of science and technology as independently evolving structures. For more direct evidence of this outlook, one need look no further than the opening sentence of Lord Zuckerman's preface:

The keystones of our century's history are those scientific and technological advances — radio, the internal combustion engine, antibiotics — which have determined the pattern of life which half the world now leads, and for which the other half strives . . . ;

or the opening paragraphs of Clark's first chapter. After evoking the political and social turmoil of late 1919, inflation and starvation in Central Europe, civil war in Russia, soldier's soviets in Britain, American rejection of the League, Clark continues:

Yet this record of battery and barricade, which added a bloody touch to the first autumn of

peace, was of infinitely less importance to the future of the human race than three events in the worlds of astronomy and physics.

More disturbing than Clark's particular choice of three events (about which more could be said) is the very idea of singling out some group of scientific and technological developments and setting them in apposition to the symptoms of vast social, economic and political upheavals in evaluating their "importance to the future of the human race". I think the evidence exists, in Clark's book and elsewhere, to sustain Raymond Williams' recent judgement:

In the case of nuclear weapons, nothing is more evident than that they were consciously sought and developed, and have continued to be consciously sought and developed. It is true that, as so often in modern technological innovations, much of the basic research had been done for quite other reasons, without foreseeing this particular result. But again as in many other comparable cases, the crucial moment of passage from scientific knowledge to technical invention, and then from technical invention to a systematic technology, depended on conscious selection and investment by an existing social order, for known and foreseen purposes (*New Left Review* 124, Nov-Dec. 1980, p.365).

Until we grasp this element of *intention* in the process of research and development in modern technology, as Williams puts it elsewhere, I fear that efforts to organize movements to control the uses of that technology will be hampered by the feeling that such developments really are inevitable. And in the case of nuclear weapons, I am not certain how much time we have left to learn the lesson. □

John Stachel is Professor of Physics at Boston University, currently editing the Einstein papers at the Institute for Advanced Study, Princeton.

More pie in the sky

Paul Davies

Gravity, Black Holes, and the Universe. By Iain Nicolson. Pp.264. ISBN 0-7153-7849-X/0-470-027111-6. (David & Charles/Halsted: 1981.) £10.95, \$24.95.

BOOKS on black holes and the Universe range from postdoctoral research treatises to the wildest distortions of science fiction and the occult. Since the explosion of interest in astronomy in general and black holes in particular, which began during the 1960s, the market has been flooded with literature on matters cosmic. How can the discerning reader tell where to go for both reliable and entertaining treatment?

Having spent many years myself on cosmological and astrophysical research, I do not envy the task of the professional writer who has to attempt to master the technical literature of a difficult and developing subject, and distil out the

features that the scientific community considers to be well established. Still more demanding is the task of grading the various conjectural topics, some of which can be bizarre, into degrees of seriousness.

Take one example: travel through a black hole. Certain solutions of Einstein's field equations display the weird property that they can be continued into regions of space and time that are forever inaccessible to us unless we fall into a black hole. These "other" spacetimes (infinite in number) are sometimes called other universes. Do they exist?

In the sense that certain idealized model spacetimes admit other universes, then the subject is a serious one, and schematic representations of these universes appear in textbooks and mathematical monographs. In spite of this, the models concerned are not supposed to apply to the real world, because they involve certain simplifications (for example, the absence of matter) that manifestly fail in reality. The prevailing opinion is that a more realistic model would not display these other universes, and that an observer who falls into a black hole will not, after all, emerge in a hitherto unknown spacetime, but will crash to oblivion in a region of unlimited tidal stress.

A writer may cover his subject thoroughly yet still miss such subtleties and proclaim that modern physics definitely predicts the existence of other universes. Certainly a glance at a textbook could give that impression. Iain Nicolson is to be congratulated that on this, and many other contentious and delicate issues, he has not only got the facts right, but has conveyed the correct degree of scepticism or reliability appropriate to each conjectural topic.

There is much more to this book than black holes, however. The author takes us gently through Kepler's laws, newtonian mechanics and special relativity, before getting on to the equivalence principle, curved spacetime and the general theory of relativity. He also applies his thoughts to the expanding Universe, and the intriguing topic of its ultimate fate.

The book is nicely illustrated with line drawings and photographs, and has a short bibliography. Although certain advanced topics, such as quantum aspects of black holes, are courageously included (and skilfully handled), the book should still appeal to general readers having only a nodding acquaintance with physics and astronomy.

My initial reaction to this book was to wonder where yet another treatment of this trendy subject could possibly fit in to an already crowded market. The answer is that it is one of the few books that maintains a reliable scientific perspective and a careful treatment, yet covers most of the "fun" topics in detail. □

Paul Davies is Professor of Theoretical Physics at the University of Newcastle upon Tyne.

The diversity of physics revealed

R.H. Dalitz

Encyclopedia of Physics. Edited by Rita G. Lerner and George L. Trigg. Pp.1,157. ISBN 0-201-04313-0. (Addison-Wesley: 1981.) \$99.50, £49.50.

THE publication of a new encyclopaedia of physics is an important event. The subject is growing rapidly, not only in accuracy of measurement and detail of understanding but also in scope. A well organized volume can quickly bring the reader up to date about developments in areas outside his own speciality where new viewpoints and techniques may have become established, and, overall, can significantly speed up the progress of physics itself.

This volume contains about 450 entries, some long (18 pages on symbols, units and nomenclature; 11 on elementary particles) and some short (a few less than a column). They range from instrumentation (particle accelerators, photomultipliers, transistors, thermocouples and so on), through classical physics (analytical dynamics, thermodynamics, acoustics, geomagnetism etc.), to quantum physics (liquid helium, atomic spectroscopy, nuclear structure, Josephson effect and so on). The articles are all by well-known physicists. They show great diversity in both objective and style, and there is considerable overlap between them.

How well will this new encyclopaedia meet our needs? Many of the major entries are really excellent — for example those on elementary particles, liquid crystals, vision and colour, and non-equilibrium thermodynamics, to mention just several. For greatest effectiveness, each major article should really provide a substantial reference list, but not all do. That on nuclear reactions, two pages long, provides an excellent classification of work in this subject but gives only one reference, whereas other articles, such as those on electronics, the Hubble effect, and relativity (special theory), give a wide range of graded references; perhaps this is overdone for solar energy, with one and a half pages of references for a six-page entry.

With a variety of viewpoints of the contributors, the multiple coverage could be a strength, but some consolidation would have obvious advantages. To give one illustration, the entries for baryons, hadrons, hyperons, and mesons, each about a page in length, could usefully have been combined into a four page article on hadrons, the other three entries each being reduced to one line, cross-referenced to hadrons. In addition, the coverage of different topics is not always in proportion to their relative importance. For example, the rather detailed contribution on weak neutral currents is given five pages, without any corresponding counterpart for weak charged currents. Also turbulence, one of the least understood areas of physics,

receives only one and a half pages.

Many of the articles read strangely for a newly published book, as if one is recalling a past time, some four or five years ago. For instance, quantum chromodynamics, our deepest and most promising theory of elementary particles, receives only one paragraph in each of the contributions on elementary particles, quantum electrodynamics, and unified field theory, despite the fact that it has provided an underlying basis for most of the theoretical work in elementary particle and high-energy physics over the past three years. The editors of an encyclopaedia face great difficulties, of course, but in this case they might have noted that a few of the articles called for some revision at the last moment.

The operational test of a reference book of this sort is whether information can be located rapidly. For this, the general index serves fairly well. I found that Higgs mechanism, radiocarbon dating, and

Peierls stress were in the index, as was Umklapp process, although with reference to p.732 where only articles on partial waves and partons are to be found. There is no entry for current algebra but the article on currents in particle theory is indexed nearby and led me to my goal. However there is no entry for Thomas precession, nor for photon multiplier, nor even for the Wigner-Eisenbud theory of the R-matrix and its application to resonance processes.

This encyclopaedia has a great deal to offer to scientists and engineers in general, as well as physicists. The underlying plan is sound and the articles generally authoritative. With so many first-rate contributions, it should have an assured place in all science libraries. It could be substantially improved, in my opinion, but, as things stand, it is the best compilation of its type available for physics today. □

R. H. Dalitz is Royal Society Research Professor in Theoretical Physics at Oxford University.

This is psychology?

John C. Marshall

Mindwatching. By H. and M. Eysenck. Pp.224. ISBN 0-7181-1937-1. (Michael Joseph: 1981.) £10.50. To be published in the USA by Doubleday.

THERE'S an old joke about how to select candidates for admission to psychology courses. The hapless individuals are asked the simple question "Why do you want to study psychology?" Those who reply "Because I want to *understand* people" are advised to read history or English literature; those who reply "Because I want to *help* people" are told to train as surgeons, farmers, garbage collectors or tax lawyers. Any other response secures admission to the psychology department.

Hans and Michael Eysenck would, I presume, disapprove. For them, the goal of professional psychologists is "to understand why people behave as they do, so that it will be possible to predict and change their behaviour"; this goal is to be achieved by controlled experimentation, the results of which are sometimes counterintuitive and thus undermine the sceptic's claim that my grandmother knew as much "psychology" as it is possible (or wise) for any mere mortal to comprehend. The Eysencks have accordingly produced this coffee-table volume in which the purported achievements, theoretical and practical, of experimental psychology are laid out for interested bystanders. Two questions therefore arise: First, to what extent are we presented with a fair summary of modern psychology? Second, are the social implications of current work outlined in a

responsible fashion?

The authors employ the stylistic device of describing in each chapter a so-called "key experiment", supplemented by anecdote and speculation. This is not in itself a bad strategy but it does lead the Eysencks into letting vaudeville prevail over rationality. For example, a section on "Attractiveness and jury decision-making" is introduced by the story of a famous nineteenth-century trial in which a woman accused of poisoning her lover was somewhat lucky to obtain a verdict of "not proven". The Eysencks speculate that since the defendant was "young, vivacious, and beautiful", these attributes may have influenced the jury and "helped to save her from the gallows". An experiment is then described in which mock juries who consider hypothetical cases of a relatively minor nature do indeed seem to be swayed a little by the physical attractiveness of the mock defendants. But the section then concludes with the claim that "most of the available research suggests that physical attractiveness has little or no effect on juries when serious crimes are involved". I can only make sense of this information by assuming the Eysencks believe that murdering one's lover is not a serious crime.

On other occasions the authors are unable to resist the temptation of a spectacular but fatally flawed study. They thus feature Rosenham's report in which a group of normal people were admitted to mental hospitals after falsely complaining that they heard voices and asking to be admitted. Following an average stay of 19

days the pseudo-patients were discharged with a diagnosis of "schizophrenia in remission". On the basis of this and a number of related pranks, Rosenham concluded that "It is clear that we cannot distinguish the sane from the insane in psychiatric hospitals". Very amusing, no doubt, but normal people who are not in the pay of psychologists either don't hear voices or, if they do, are not sufficiently worried by it to seek admission to hospital. The Eysencks themselves quote Seymour Kety's withering comment — "If I were to drink a quart of blood and, concealing what I had done, come to the emergency room of any hospital vomiting blood, the behaviour of the staff would be quite predictable. If they labelled and treated me as having a bleeding peptic ulcer, I doubt that I could argue convincingly that medical science does not know how to diagnose that condition" — but they are not thereby inhibited from describing Rosenham's jest as a "key experiment".

Not all of the selections are as silly as Rosenham's study. The book contains quite sensible sections on eye-witness testimony, divided attention and forgetting. Similarly, the chapter on attempts to teach sign language to chimpanzees is an improvement on many popular treatments of the topic. But these islands of sanity are lost in the surrounding sludge of trivia and titillation. What can one say of a book which proclaims that psychology should "be regarded as the most important of all the sciences" and then reports that "Small breasts were preferred by men with fundamentalist religious beliefs suffering from mild depression"?

There is little improvement when the Eysencks draw out the practical implications of their choice of topics. We are informed that a process of conditioning can turn our children into "well-integrated, honest members of society" and referred to Chapter 6 for the experimental evidence that this is so. Chapter 6, however, reports that obsessive and compulsive disorders can be treated by "behaviour therapy". In Chapter 22, a discussion of Hans Eysenck's "theory" of personality is held to support the claim that "The egalitarian principle underlying socialism is in conflict with biological reality and can only be enforced by dictatorship". Exactly how this conclusion is reached remains shrouded in mystery. An appallingly ill-controlled "experiment" on the effects of two types of play-group prefaces the remark that "the advocates of progressive teaching . . . have turned our pre-school groups, kindergartens and primary schools into permissive places of no work and all play, from which all socializing and restraining influences have been banned".

One can only hope that this grubby little opus will sink without trace. It totally fails to represent those aspects of psychology where results of some intellectual depth have been obtained in the study of perception, language and skilled action; it similarly ignores many areas of occupational, clinical and educational research where careful, thoughtful work has served to enhance the quality of life. □

John C. Marshall is in the Neuropsychology Unit, part of the Neuroscience Group at the Radcliffe Infirmary, Oxford.

Check-list of Birds of the World. For example, Terres leaves the titmouse family intact and includes the long-tailed tits (Aegithalinae) and penduline tits (Remizinae) as subfamilies rather than treating them as distinct families as has been recently practised. On the other hand, the thrushes are retained in their own family, Turdidae, and there is no mention of their current placement as a subfamily in the Old World flycatcher family, Muscicapidae.

While the book has many virtues it falls short of the stated goal of providing information on all the birds that have nested or been sighted in Alaska, Canada, Greenland, Bermuda, Baja California and the 48 contiguous states of the USA. Omissions are conspicuous, for example, in the Alaskan avifauna. While the 1978 update of the status of Alaskan birds is referenced in numerous places, and 1980 Alaskan references are also given, it is annoying that no less than 22 Alaskan species given in the 1978 publication are ignored. For instance, while the willow warbler (an Old World warbler) is discussed, nothing is said about its North American status. Only its Eurasian distribution is mentioned, so readers are not aware that it has occurred as an accidental in Alaska.

Another flaw is that the book's attractiveness is compromised by lumping colour plates together in sections. For example, located in the middle of the thrush family species accounts are 31 pages of colour plates of tanagers, thrush family, titmouse family, trogon, tropicbirds, tropical (American blackbird) family, turkey, vireo family and New World vultures and condors — rather a potpourri even though they are in alphabetical order. As with most undertakings of this magnitude other errors, mainly of editorial style, creep in. Under "Fecal Sac" we are referred to "Nests and Nestlings" but there is nothing on the sac there. A second reference under "Fecal Sac" is "Nest Sanitation" which is there, but under that title we are referred further to "Young and Their Care" which contains a discussion of nest sanitation and fecal sacs. Further minor inconsistencies are the listing of "Courtship" with sub-topics such as *courtship rituals* but not *courtship feeding*. This latter appears as a completely separate topic followed by "Courtship Flight", which then only refers the reader back again to "Courtship" and "Flight".

Despite my less-than-total enthusiasm with the *Encyclopedia*, on balance Terres has done a remarkable job in producing a scientifically accurate tome relatively free of typographical errors. With over 7 million bird-watchers in the US alone (as of 1970), of which at least 3 million were considered to be seriously interested in field ornithology, there should be a receptive audience to savour this book. □

Clayton M. White is a Professor of Zoology at Brigham Young University, Provo, Utah.

Find the (North American) bird

Clayton M. White

The Audubon Society Encyclopedia of North American Birds. By John K. Terres. Pp.1,109. ISBN 0-3944-6651-9. (Knopf: 1980.) \$60 US only.

IN 1896 Professor Alfred Newton published his *Dictionary of Birds*. Thus a mould was cast. But it took another 68 years, until 1964, through the British Ornithologists' Union and under the editorship of A. Landsborough Thomson, for another such work to appear: *A New Dictionary of Birds*. Now, only 16 years later, another giant encyclopaedia is on the market. To me, this current volume appears to be an attempt to update a considerable body of material from the now elusive *New Dictionary*. Its scope, however, is mainly restricted to North America.

The book is designed to be a single-volume work for both layman and professional, but I suspect the lay will benefit most. There are nearly 6,000 alphabetically arranged entries and cross references, illustrated with 875 full-colour

photographs of bird species and over 800 black-and-white drawings. The scope of the entries is impressive: Birdhouse, Bird Lice, Birds as Collectors, Birds as Watchdogs, Bird Skin, Bird-Watcher, Bisexuality, Bittern, Blackburne, among the "B's" for example. Abbreviated life histories of 847 birds are provided. For each species are given a description of the Latin or Greek origin of the scientific name (a most valuable feature), the bird's geographical range, colour and distinctive morphological features, habitat, feeding habits, other names, nest, eggs, and other features of interest (such as who parasitizes whom). Terres cites over 4,000 references, ranging in date from the turn of the 1800s to 1980, all of which are listed in a bibliography. The bulk of several alphabetical sections is taken up by species accounts. The taxonomic treatment is fairly current by both European and American standards, at least in common names, but is more conservative and departs somewhat from the nomenclature of Peter's

REVIEW ARTICLE

Localization and electronic properties in amorphous semiconductors

J. Mort

Xerox Webster Research Center, Webster, New York 14580, USA

J. Knights

Xerox Palo Alto Research Center, Palo Alto, California 94304, USA

The transport and photogeneration of different classes of amorphous semiconductors are considered. These properties are affected by localization associated with long-range disorder and defects which in turn reflect the underlying chemical nature of each class.

PRESENT interest in the electronic properties of amorphous semiconductors is the result of several related factors. The demonstration^{1,2} that amorphous Si, (a-Si), like the crystalline form, can be doped to form either a *n* or *p* extrinsic semiconductor suggested new ways of using amorphous materials as electronic materials.³⁻⁶ Other classes of amorphous solids than this tetrahedral material, have been extensively studied in recent years, including chalcogenides such as amorphous selenium (a-Se), amorphous arsenic triselenide, (a-As₂Se₃) and amorphous molecular or organic solids⁷⁻¹⁰. It is now clear that although amorphous solids have certain characteristic features in common, there are also distinctive differences. Our purpose here is to compare and contrast the electronic properties of these three groups of amorphous solids—the chalcogenides, the tetrahedrally bonded materials and molecular systems.

The most obvious consequence of disorder in amorphous solids is that intrinsic localized states exist within the energy gap. The experimental and theoretical problems of understanding the effect of disorder on the photoelectronic properties of solids are considerable, but much progress has been made in the past decade in three directions. Light has been thrown on the origin and nature of localized states, and their relationship with the specific chemical nature and the bonding of the constituent building-blocks. The mechanism of charge transport and photogeneration has been partially understood; quantitative studies of the consequences of doping have begun. In what follows, these three areas will be discussed for the three general classes of amorphous solids.

Nature of localized states

In the development of crystalline semiconductors, major conceptual advances such as the understanding of band-gaps arose from the recognition of the long-range periodicity and symmetrical arrangement of the atomic building blocks—intrinsic features which are essentially independent of the method of preparation. Most technical applications of crystalline semiconductors and their associated physics are, however, connected with deviations from idealized order. The localized states within the band-gap produced by defects, impurities, surfaces and interfaces control many properties, particularly those involving carrier transport and generation-recombination processes; such perturbations are strongly influenced by both the method of preparation and subsequent treatment.

Viewed in this way amorphous semiconductors exhibit remarkably similar features. The existence of band gaps in the absence of long-range periodicity has a firm theoretical basis in tight-binding approaches. It has been predicted^{11,12}, based on Anderson's work on localization¹³, that even in the absence of defects the potentials fluctuating because of the disorder would result in localized tail states at the conduction and valence band edges. Figure 1 shows that the density of states distribution in an amorphous solid can be divided into three regions. For energies greater than ϵ_c and less than ϵ_v , electronic states are extended or band-like in character. For energies close to these values, the scattering length of the electron wave function approaches interatomic distances. Conduction then occurs as a Brownian diffusive motion with estimated mobilities between 1 and

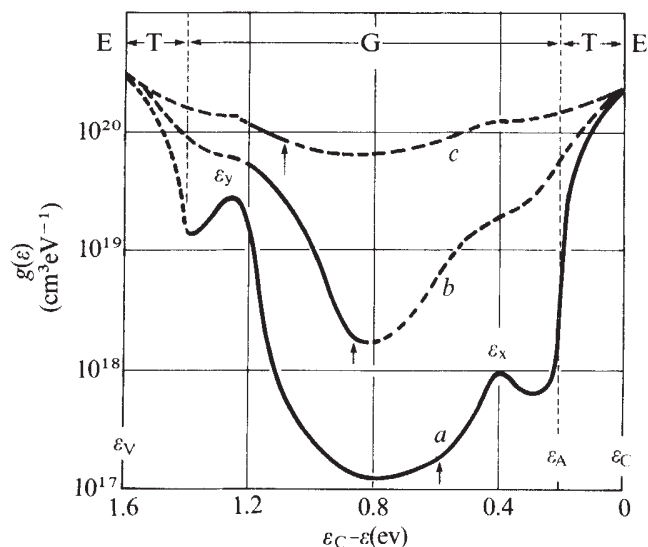


Fig. 1 Density-of-state distributions for a-Si specimens. *a*, Glow discharge specimen, $T_d \approx 520$ K; *b*, glow discharge specimen, $T_d \approx 350$ K; *c*, evaporated specimen. The solid lines indicate results obtained from field-effect experiments and the arrow on each curve shows the position of the Fermi level. E, extended states; T, tail states; G, gap states. (After ref. 15.)

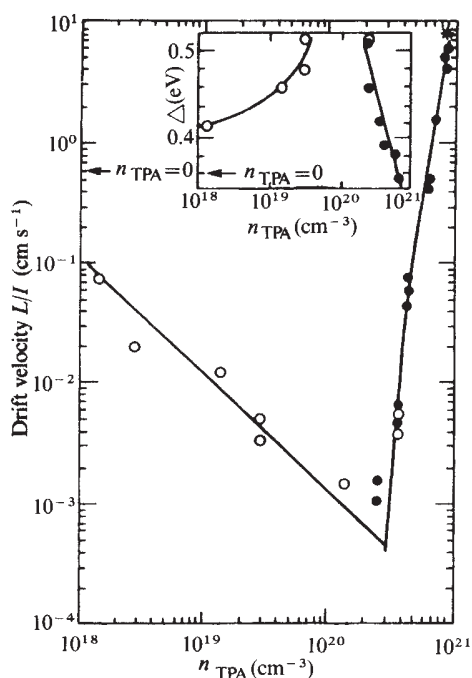


Fig. 2 Concentration dependence of the drift velocity of holes in a mixed doped polymer showing the trap-controlled hopping process. NIPC concentrations: $\circ$, $1 \times 10^{21} \text{ cm}^{-3}$; $\bullet$, 0; $*$, $2 \times 10^{20} \text{ cm}^{-3}$. The points correspond to samples with varying concentrations of TPA, which has a lower ionization potential than NIPC. (After ref. 25.)

$10 \text{ cm}^2 \text{ V s}^{-1}$ (ref. 14). Just below ε_c and above ε_v , there exist the tail states due to Anderson localization that are associated with satisfied but distorted bonds. Anderson localization begins when U_0/J (where U_0 is the range of site energy disorder and J is the transfer integral between nearest neighbours) exceeds a critical value at which an electron at a given site at temperature $T = 0$ becomes localized, that is, it can no longer diffuse away. Defects arising from incomplete or dangling bonds or impurities give rise to the deep lying gap states. As in the crystalline case, the nature and density of these states depend on conditions of preparation and treatment.

The determination of detailed density of state distributions has proved difficult. In the case of amorphous Si some information has come from studies of the field-effect¹⁵ as shown in Fig. 1. Although quantitative details and the relative contributions of surface and bulk states are uncertain, the general features are consistent with independent studies of conductivity, thermopower and drift mobility. The distribution and number of localized states in a-Si:H prepared by glow discharge of silane may be controlled by the preparation conditions (Fig. 1). This is partly because hydrogen satisfies dangling silicon bonds at defects in the extended amorphous network. Many of the associated electronic defect states are thereby removed from the forbidden gap, which suggested that impurity doping above the residual levels might be possible^{1,2}.

These features are general and may be applied to the covalently bonded chalcogenides. However, all efforts to observe a significant field-effect have failed¹⁶. This is surprising¹⁷ because the efficiency of amorphous Se in electrophotography indicates a substantial carrier range of $\sim 50 \mu\text{m}$, which implies gap state densities $\sim 10^{14} \text{ cm}^{-3}$. Recent intrinsic defect models for gap states, which take account of the specific atomic bonding of chalcogen atoms coupled with the law of mass action, may resolve this conflict^{18,19}. The valence alternation pair defects (VAPS) lead to a thermodynamic equilibrium of D^+ and D^- states, which effectively make the Fermi level independent of

the actual density of the state; in other words the Fermi level cannot be moved relative to the bands either by an applied field as in the field effect experiment or doping.

In disordered molecular solids, the building blocks are molecules. Even in molecular crystals the weak Van der Waal's forces between molecules lead at best to very narrow energy bands. In this case, the disorder, associated with pendant group polymers such as poly(*N*-vinylcarbazole) or molecularly-doped polymers means that extended states cannot be involved⁸. Such systems, particularly molecularly-doped polymers, therefore must consist of purely localized states which are associated with the dopant molecules⁸, in this way they are almost ideal for studying phenomena associated with localized states.

Electronic transport

The most unambiguous information on electronic transport in amorphous materials has come from measurements of the drift mobility by the time-of-flight technique. In this method the time, t_T , for a sharply defined spatially and temporally photoexcited charge sheet to cross a sample of known thickness L , with an applied bias field E , is measured. Because $t_T = L/\mu E$, the carrier mobility μ can be directly determined. Table 1 gives typical values of the drift mobilities for both electrons and holes in representative amorphous materials. In all cases the mobilities are unity or less and are thermally activated. There are various explanations for the nature of this electronic transport. No conduction can occur in the localized tail and gap states at $T = 0$. At finite temperatures, however, thermally activated hopping transport between localized states can occur and the mobility $\mu(\varepsilon)$ at an energy ε can be written²

$$\mu(\varepsilon) \approx \frac{1}{6} \frac{eR^2(\varepsilon)}{kT} \nu_{ph} \exp(-2\alpha R) \exp(-W/kT) \quad (1)$$

where R is the average distance between sites at energy ε and is determined by the density-of-state distribution. The term $\exp(-2\alpha R)$ describes the overlap of the wave functions on neighbouring hopping sites and α^{-1} is a measure of the spatial decay of the wave function^{2,7}; $\nu_{ph} \exp(-W/kT)$ represents the probability that the localized carrier hops to a site with energy higher by W . This transport mechanism solely controlled by hopping from one localized state to another should be observed in amorphous materials. However, before doping studies proved possible, this mechanism could only be deduced with some ambiguity from the combined observation of low mobility values and thermal activation as these features are also characteristic of transport involving extended state transport modulated by trapping in energetically shallow traps²⁰.

Direct observation of pure hopping between localized states became possible with the advent of molecularly-doped polymers²¹. These are otherwise electrically-inert polymers into which molecules can be dispersed at high controllable concentrations $< 10^{21} \text{ cm}^{-3}$ and constitute a random ensemble of molecules which have essentially no mutual interaction. As a result they consist of pure localized states without the complication (and advantage) of extended states. Measurements of the effective mobility by the time-of-flight technique as a function of

Table 1 Representative hole (electron) drift mobilities at room temperature for examples of different types of amorphous materials

	$(\text{cm}^2 \text{ V}^{-1} \text{ s}^{-1})$	$(\text{cm}^2 \text{ V}^{-1} \text{ s}^{-1})$
a:Se	$(5-8) \times 10^{-3}$	$0.11-0.19$
a:As ₂ Se ₃	—	$\sim 10^{-5}$
a-Si:H	0.1	5×10^{-4}
PVK	—	$< 10^{-6}$

See refs 2, 8, 27.

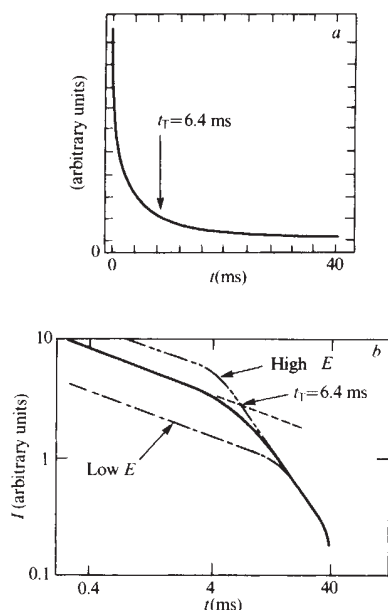


Fig. 3 *a*, Hole transit pulse in a-As₂Se₃ showing current (I) versus time t_T in linear scales. The transit time t_T deduced from *b* is indicated. *b*, Hole transit pulses in a-As₂Se₃ in logarithmic scales. The middle trace is the same pulse shown in *a*. The sharp change in slope occurs when the leading edge of the charge sheet has transited the sample. This transit time t_T is defined by the point of intersection of the extrapolated linear parts of the trace. The upper and lower current show the shift in t_T as the field E is changed.

the average intermolecular separation R have been made for the transport of holes in polycarbonate films (Lexan) doped with a molecule tri-*p*-tolylamine (TTA). The highest mobility is $\sim 10^{-5} \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ giving a value of $\alpha^{-1} \sim 1.1 \text{ \AA}$ (ref. 22). A similar transport study has been reported for a-Si:H which was doped with phosphorus (donor) impurities and compensating boron (acceptor) impurities to produce samples with various concentrations of compensated and therefore unoccupied donor states. By contrast hopping between such donor states yields values of $\alpha^{-1} \sim 22 \text{ \AA}$ (ref. 23).

The difference in localization parameter α^{-1} shows that compared with covalent amorphous solids charge in amorphous molecular solids is highly localized. This reflects the considerable internal structure of a molecular localized state, with intramolecular vibrational modes so that strong electron-phonon coupling can result in small polarons²⁴. These features lead to the exceptionally low mobilities found in these systems.

A transport process in molecularly-doped polymers which is also expected to occur in amorphous solids is trap-controlled hopping. In this mechanism a carrier can temporarily reside, between hops, in a trap, thus reducing the average drift mobility and introducing an additional activation energy due to trap depth. Figure 2 shows the results of trap hopping by admixtures of different molecules in a polymer matrix. These experiments²⁵ use controlled small quantities of a molecule, triphenylamine (TPA), with a lower ionization potential than a second molecule *N*-isopropylcarbazole (NIPC) present in very high concentrations. The holes normally propagate by hopping between the closely spaced NIPC molecules, but the introduction of TPA results in the trapping of a hole and its temporary removal from the transport channel. As the concentration of TPA and number of trapping events rises, the average carrier velocity falls. A concentration of TPA is ultimately reached at which (because of sufficient overlap between TPA molecules) TPA becomes an alternative and preferred transport channel, so the carrier velocity rises again.

Despite detailed studies of mobilities in the chalcogenide glasses the transport mechanism is not as clear⁷, although there is fairly strong evidence for trap-controlled extended state motion in a-Se, and trap-controlled hopping in a-As₂Se₃. An idealized current transit pulse should be rectangular in shape in the time-of-flight technique. Actual transit pulses in amorphous solids rarely have this character (an exception being amorphous Se at room temperature²⁶) and in some cases give no direct indication of transit time²⁷. This dispersive transport has been observed in all three groups of amorphous solids. This general phenomenon can be explained by the theory of transport in disordered systems and occurs because of a broad distribution of transition probabilities between sites due to the inequivalence of site energies within the disordered solid²⁸. This model predicts that a logarithmic plot of transit current versus time should alter when the leading edge of the charge sheet completes its transit.

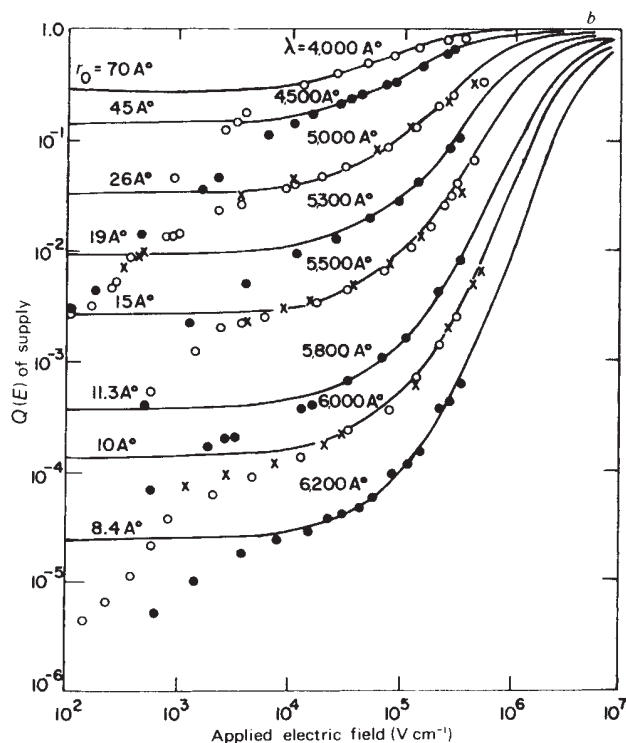
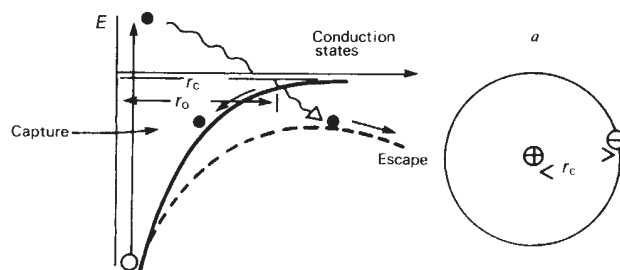


Fig. 4 *a*, Dissociation of bound electron hole pair. r_c is the distance at which this Coulomb binding energy is comparable with kT . Excess energy of the carrier from the excitation is lost by thermalization processes which occurs at a thermalization length r_0 . *b*, Solid lines are the quantum efficiencies predicted by the Onsager theory for $\Phi_0 = 1$ and the dielectric constant of a-Se ($\epsilon = 6.0$) for the thermalization distances r_0 indicated. The circles and crosses show the experimental quantum efficiency for the photoinjection of holes in a-Se versus applied field for different exciting wavelengths. Included are data for films of two different thickness: $\bullet$, $\circ$, $L = 44.0 \mu\text{m}$; $\times$, $L = 3.4 \mu\text{m}$. (After ref. 30.)

Table 2 Amorphous materials are different and may be categorized according to how the conductivity ($\sigma = ne\mu$) is affected by doping

Chalcogenides	Tetrahedral	Molecularly doped polymers
σ , relatively weak dependence $\Delta\sigma$ related to $\Delta\mu$ (Fermi level immobile)	σ strong dependence $\Delta\sigma$ related to Δn (Fermi level moves)	σ strong dependence $\Delta\sigma$ related to Δn and $\Delta\mu$

An example of such dispersive transits in a-As₂Se₃ are shown in Fig. 3. In Fig. 3a a transit signal, showing the photoinduced transient current versus time in linear scales, is essentially featureless. The same transit signal in Fig. 3b is displayed with logarithmic current and time scales. The predicted change in slope when the leading carriers have transited the sample is indicated at t_T which shifts to shorter times at higher fields, E . The concept of mobility now loses its meaning as no average velocity of the charge packet is definable, and the predicted thickness-dependent mobilities have been observed. This type of dispersive phenomena appears in recombination processes, time dependent luminescent decay and trapping²⁹.

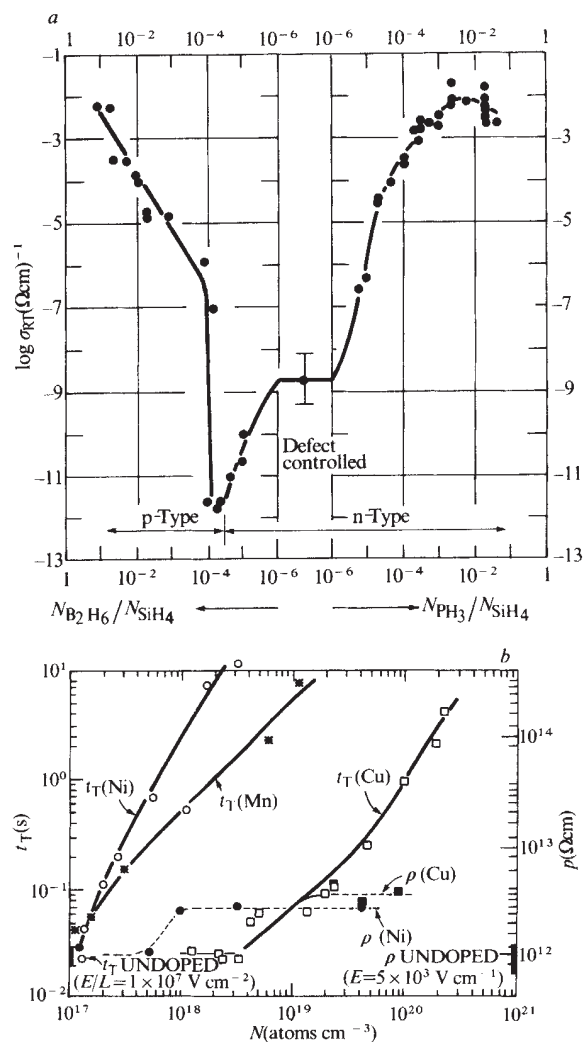


Fig. 5 a, Room-temperature conductivity σ_{RT} of n and p-type amorphous Si specimens as a function of the gaseous impurity ratio. For the right-hand curve, this is the number of phosphine to silane molecules in the gas mixture used for specimen preparation. On the left it is the corresponding diborane to silane ratio. The centre refers to undoped specimens. (After ref. 1.) b, log-log plot of transit time and resistivity ρ versus concentration N of: $\circ$, $\bullet$, Ni; $\square$, $\blacksquare$, Cu doped a-As₂Se₃. (After ref. 37.)

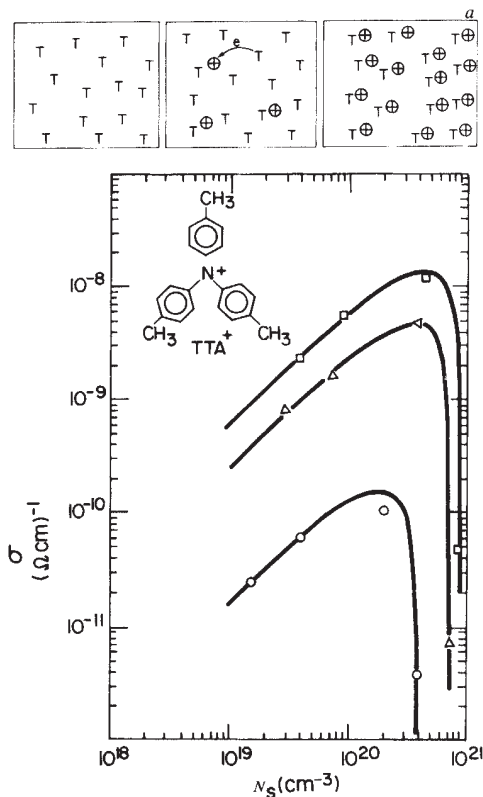


Fig. 6 a, Chemical control of conductivity in a molecularly-doped inert polymer matrix. Fractional oxidation of neutral transport molecules T is achieved by controlled addition of an oxidizing agent SbCl₅. The oxidized transport molecules are radical cations T^{•+} which constitute free carriers. Progressive oxidation of the neutral molecules leads to increasing free carrier concentrations at the expense of the neutral molecules necessary for the transport mechanism. b, Dependence of conductivity on concentration of oxidized transport molecules N_S . The molecule tri-p-tylamine (TTA) is dispersed in Lexan[®] polycarbonate. The total molecule concentrations were: $\square$, $8.6 \times 10^{20} \text{ cm}^{-3}$; $\triangle$, $7.2 \times 10^{20} \text{ cm}^{-3}$; $\circ$, $0.86 \times 10^{20} \text{ cm}^{-3}$. (After ref. 40.)

Photogeneration

Photogeneration is a critical process in any photoelectronic material. The way in which absorbed photons produce photogenerated carriers is affected by disorder within the material. Geminate recombination is likely to be a limiting step in the photogeneration process in materials where the mean free paths of photoexcited carriers are characteristically low³⁰—a natural condition of amorphous materials. Figure 4a illustrates this photogeneration process whereby the initial photoexcitation leads to electron-hole pairs, which diffuse within their mutual Coulomb well. There is a finite probability, which decreases with applied field, that the pair will combine at the initial excitation site rather than dissociate into a free electron and a hole. Three critical parameters are; Φ_0 , the efficiency of production of thermalized electron-hole pairs per absorbed photon; r_0 , a characteristic thermalization length (related to the mean free path); and r_c , the coulomb radius or the distance at which the coulomb energy $\sim kT$. Precise measurements³⁰ for a-Se indicate Φ_0 (the maximum possible quantum efficiency) is unity and there is evidence that this also applies to a-Si (refs 31, 42) and a-As₂Se₃ (ref. 32). For disordered organic systems, however, Φ_0 can be ~ 0.1 or even smaller. This reflects the existence of intramolecular relaxation processes, radiative or non-radiative, which compete with the photogeneration or dissociation of the photoexcited molecule³³. Figure 4b shows a treatment of this problem due to Onsager³⁴ for a material with $\Phi_0 = 1$ and the dielectric constant of a-Se ($\epsilon = 6.0$). The various curves are for

different assumed values of r_0 . For a given value of r_0 the quantum efficiency remains field independent up to $\sim 10^4$ V cm $^{-1}$ after which it becomes field dependent and approaches unity in a sufficiently high field. As r_0 increases the initial or zero field quantum efficiency increases and the subsequent field dependence decreases. Where a curve for a particular value of r_0 lies on the absolute quantum efficiency axis depends on the coulomb radius of the material which is determined by the dielectric constant. Typical coulomb radii are ~ 49 Å for a-Si, ~ 100 Å for a-Se and ~ 250 Å for most organics, this would indicate descending order for the photogeneration efficiency for comparable values of r_0 and Φ_0 .

Experimental data for a-Se shown in Fig. 4b explain the macroscopic observation that a-Se is more panchromatic in absorption than in photosensitivity³⁰. This arises because r_0 being wavelength dependent ranges from 70 Å at a wavelength 4,000 Å to 10 Å at a wavelength of 6,000 Å. Recent evidence shows that photogeneration in a-Si:H and a-As₂Se₃ may also be geminate^{31,32}. Unlike the case of a-Se, however, r_0 has at most a very weak wavelength dependence. A field-dependent quantum efficiency has implications for ultimate solar cell conversion efficiencies, this is because of the spatial field distribution within the barrier of Schottky or p-n cells which, with the absorption profile of the light, leads to photons being absorbed in a range of fields and effectively leads to a spatial distribution of quantum efficiencies³⁵.

Doping of a-semiconductors

One of the most important advances in the field of amorphous materials has been the demonstration that n or p type a-Si:H and a-Ge:H could be produced by appropriate doping; there is no evidence that this applies to chalcogenides, although it is possible in a-C:H (ref. 36). It seems possible to group amorphous materials according to the way in which doping affects their electrical conductivity σ , as indicated in Table 2. Detailed studies of doping in amorphous Si and, to a lesser degree, in a-carbon indicate that σ is strongly dependent on doping. Conductivity, thermopower and drift mobility measurements demonstrate that this is associated with a shift in the Fermi level which affects both the type and density of carriers controlling the conductivity (see Fig. 5a)². On the other hand, relatively small changes in the transport occur, except for special cases such as inter-donor hopping in compensated samples. By contrast, extensive doping studies of chalcogenides, for doping levels well below the alloy range, do not result in significant movement of the Fermi level or change in conductivity³⁷. However, substantial and systematic changes in the drift mobility of holes have been reported in a-As₂Se₃. Because the drift mobility can be the microscopic mobility modified by trapping, the fact that these changes in the drift mobility are not reflected in the conductivity (see Fig. 5b) indicates that the impurities are changing the density of traps which controls the mobility³⁷. This

insensitivity of the position of the Fermi level in chalcogenides to doping is consistent with the absence of a field effect and with other defect models^{18,19} based on valence alternation pair defects, D⁺ and D⁻, which are positively and negatively charged defects whose creation is governed by the law of mass action. Such defects produced under thermal equilibrium effectively pin the Fermi level. It is possible that doping of chalcogenides in non-equilibrium conditions could result in an unpinning of the Fermi level and production of extrinsic semiconductors. Whether this can be done in a stable manner is unclear.

Recent studies confirm that the densities of defects can be affected by heating³⁸. Thus photoelectronic properties of chalcogenides immediately after preparation are dependent on thermal history. Characteristic equilibrium behaviour, independent of the method of preparation is, however, achieved after annealing^{38,39}. The latter are consistent with Spear's observation that the most remarkable feature of chalcogenides is the reproducibility of their mobility values, a feature probably unmatched by even the best crystalline semiconductors¹⁶. This then, seems to be related to the ability of chalcogenide glasses to self-anneal defects efficiently.

In the disordered organic materials, doping is accomplished by the controlled chemical oxidation of some of the dopant molecules⁴⁰. These radical cations are the equivalent of 'free' carriers because electrons can hop from neighbouring neutral molecules and thereby the 'free' hole propagates (see Fig. 6a). Thus at low oxidation levels the conductivity rises as the 'free'-carrier concentration rises. However, because no extended states exist and free carriers are highly localized and associated with specific molecules, such free carriers are created at the expense of transport paths. In the extreme case, where the total oxidation has occurred, none of the neutral molecules necessary for the transport process remain and the conductivity must go through a maximum as indicated in Fig. 6b (ref. 40). It is possible to produce films containing $\sim 10^{20}$ cations cm $^{-3}$ which form what Mott has termed a degenerate polaron gas⁴¹. In this case, therefore, the conductivity is controlled by changes in both free carrier concentrations and mobility where these are not independent.

Conclusions

We have outlined the transport and photogeneration properties in three general groups of amorphous materials which make up most of materials known as amorphous semiconductors. In particular the different effects of localization of electronic wavefunctions due to long range disorder and those due to defects have been considered. The picture that emerges is that although features characteristic of the disordered state do exist, amorphous materials exhibit a rich diversity of behaviour which reflects their underlying chemical composition. In this sense the similarities between disordered and crystalline solids are closer than was originally perceived.

1. Spear, W. E. LeComber, P. G. *Phil Mag.* **33**, 935 (1976).
2. Spear, W. E. *Adv. Phys.* **26**, 811 (1977).
3. Carlson, D. E. & Wronski, C. R. *Appl. Phys. Lett.* **28**, 671 (1976).
4. Spear, W. E., LeComber, P. G., Kinmond, S. & Brodsky, M. H. *Appl. Phys. Lett.* **28**, 105 (1976).
5. Gibson, R. A., Spear, W. E., LeComber, P. G. & Snell, A. J. *J. Non-Cryst. Solids* **35-36**, 725 (1980).
6. LeComber, P. G., Spear, W. E. & Garth, A. *Electron. Lett.* **15**, 179 (1979).
7. Pfister, G. *Phil. Mag.* **36**, 1147 (1977).
8. Mort, J. & Pfister, G. *Polymer-Plast. Technol.* **12**, 89 (1979).
9. Mort, J. *Adv. Phys.* **29**, 367 (1980).
10. Mort, J. *Science* **208**, 819 (1980).
11. Mott, N. F., *Electronic and Structural Properties of Amorphous Semiconductors* (eds LeComber, P. G. & Mort, J.) (Academic, London, 1973).
12. Cohen, M. H., Fritzsche, H. & Ovshinsky, S. R. *Phys. Rev. Lett.* **22**, 1065 (1969).
13. Anderson, P. W. *Phys. Rev.* **109**, 1492 (1958).
14. Cohen, M. H. *J. Non-Cryst. Solids* **4**, 391 (1970).
15. Madan, A., LeComber, P. G. & Spear, W. E. *J. Non-Cryst. Solids* **20**, 239 (1976).
16. Spear, W. E. *Amorphous and Liquid Semiconductors* (eds Stuke, J. & Brenig, W.) (Taylor and Francis, London, 1974).
17. Nielsen, P. *Phys. Rev.* **6**, 3739 (1972).
18. Mott, N. F., Davis, E. A. & Street, R. A. *Phil. Mag.* **32**, 961 (1975).
19. Kastner, M., Adler, D. & Fritzsche, H. *Phys. Rev. Lett.* **37**, 1504 (1978).
20. Dolezalek, F. *Photoconductivity and Related Phenomena* (eds Mort, J. & Pai, D.) (Elsevier, Amsterdam, 1975).
21. Mort, J., Pfister, G. & Grammatica, S. *Solid St. Commun.* **18**, 693 (1976).
22. Mort, J., Grammatica, S., Sandman, D. J. & Troup, A. J. *electron. Mater.* **9**, 411 (1980).
23. Allan, D., LeComber, P. G., & Spear, W. E. *Proc. 7th Int. Conf. on Amorphous and Liquid Semiconductors* (ed. Spear, W. E.) 323 (CICL, University of Edinburgh, 1977).
24. Emin, D. *Electronic and Structural Properties of Amorphous Semiconductors* (eds LeComber, P. G. & Mort, J.) (Academic, London, 1973).
25. Pfister, G., Mort, J. & Grammatica, S. *Phys. Rev. Lett.* **37**, 1360 (1976).
26. Pfister, G. *Phys. Rev. Lett.* **36**, 271 (1976).
27. Pfister, G. & Scher, H. *Adv. Phys.* **27**, 747 (1978).
28. Scher, H. & Montroll, E. W. *Phys. Rev.* **B12**, 2455 (1975).
29. Vardeny, Z., O'Connor, P., Ray, S. & Tauc, J. *Phys. Rev. Lett.* **84**, 1267 (1980).
30. Pai, D. M., & Enck, R. C. *Phys. Rev.* **B11**, 5163 (1975).
31. Williams, R. & Crandall, R. S. *RCA Rev.* **40**, 371 (1979).
32. Enck, R. C. & Pfister, G. *Photoconductivity and Related Phenomena* (eds Mort, J. & Pai, D.) Ch. 7 (Elsevier, Amsterdam, 1975).
33. Inokuchi, H. *Photoconductivity and Related Phenomena* (eds Mort, J. & Pai, D.) Ch. 5 (Elsevier, Amsterdam, 1975).
34. Onsager, L. *Phys. Rev.* **54**, 554 (1938).
35. Chen, I. & Mort, J. *Appl. Phys. Lett.* **37**, 952 (1981).
36. Meyerson, B. & Smith, F. W. *Solid St. Commun.* **34**, 531 (1980).
37. Pfister, G. & Morgan, M. *Phil. Mag.* **B41**, 209 (1980).
38. Abkowitz, M. & Enck, R. C. *J. Non-Cryst. Solids* **35-36**, 831 (1980).
39. Abkowitz, M. *The Physics of Selenium and Tellurium* (eds Gerlach, E. & Grosse, P.) (Springer, Berlin, 1979).
40. Troup, A., Mort, J., Grammatica, S. & Sandman, D. J. *J. Non-Cryst. Solids* **35-36**, 151 (1980).
41. Mott, N. F. *J. Non-Cryst. Solids* **35-36**, 1321 (1980).
42. Mort, J. *et al. Appl. Phys. Lett.* **38**, 277 (1981).

ARTICLES

Resonant structure of the asteroid belt

Stanley F. Dermott & Carl D. Murray

Center for Radiophysics and Space Research, Space Sciences Building, Cornell University, Ithaca, New York 14853, USA

Statistical techniques are applied to asteroid orbital data in an attempt to define the characteristics of the Kirkwood gaps in the distribution of the asteroids. A significant tendency is found for the eccentricities and inclinations of asteroid orbits to increase away from the gaps. These effects are observed throughout the distribution of main-belt asteroids. They are not merely confined to the immediate vicinities of the resonances and are not caused by observational selection effects. There is no evidence that the magnitude-frequency distribution of the asteroids changes near the gaps. Evidence is given for the existence of gaps in Hirayama families, possibly of a different nature from the classical Kirkwood gaps at locations corresponding to high order commensurabilities with Jupiter. These results support the theory that the Kirkwood gaps were formed by gravitational processes acting on individual asteroids throughout their lifetimes.

ALTHOUGH gaps in the distribution of the asteroid population are generally accepted to be associated with resonances with Jupiter¹, the exact mechanism which produces the gaps remains in dispute. The four major hypotheses² proposed to explain these Kirkwood gaps¹ are: (1) the statistical hypothesis (the proposal that asteroids are librating about the gaps and are therefore only rarely seen crossing the exact resonances), (2) the gravitational hypothesis (the gaps are formed by purely gravitational interactions with Jupiter), (3) the collisional hypothesis (the gravitational effects of Jupiter increase collision rates within the gaps, causing a wearing down or removal of material) and (4) the cosmogonic hypothesis (the gaps represent regions where asteroids failed to form during the early history of the Solar System). Although there are problems associated with each hypothesis, recent work³⁻⁶ has provided convincing evidence against the statistical hypothesis. It is important, therefore, to recognize which properties of the Kirkwood gaps could help to distinguish between the gravitational, collisional and cosmogonic hypotheses. Little work has been carried out in this field partly because of the failure to apply established statistical methods to the wealth of data available.

Resonance

The integral with respect to time of the expansion of the disturbing function for the gravitational interaction between two bodies (see, for example, ref. 7) contains divisors which become small when the ratio of the mean motions is close to a simple fraction of the form $p/(p+q)$, where p and q are integers. Such locations in the asteroid belt correspond to resonances with Jupiter and q is called the order of the resonance. The strength of a given resonance depends on several factors which include the distance between the bodies and the value of q : a large value of q implies a weak resonance. In the present study, we include all resonances with $q \leq 6$ which lie between the 1/3 and 1/2 resonances with Jupiter. These are the 1/3, 3/8, 2/5, 3/7, 4/9, 5/11 and 1/2 resonances. The limit $q \leq 6$ is arbitrary but reasonable. If Kirkwood gaps do have any characteristics other than a lack of asteroids, one would expect them to be more prominent in the vicinity of the strongest resonances.

Asteroids outside the 1/2 resonance are omitted. The number of asteroids between this resonance and Jupiter is small and some resonances in this region are associated with groupings of asteroids rather than gaps (see ref. 6). Asteroids inside the 1/3 resonance are also omitted because their orbits may have been influenced by Mars as well as by Jupiter⁸. Furthermore, the

resonances in this region, which is distant from Jupiter, are weak.

The *Tucson Revised Index of Asteroid Data* (TRIAD) is a convenient source of information on the orbital elements and physical properties of asteroids. In particular, Williams⁹ has calculated sets of proper elements (orbital elements after the effects of short-period oscillations have been removed) for 1,687 asteroids. Hirayama¹⁰ recognized certain groupings in the distribution of proper elements which he called families. The inclusion in our study of large numbers of asteroids whose orbital elements are similar or correlated for reasons other than those associated with resonance would give misleading results. We therefore exclude any asteroid which is designated a family member in the TRIAD file.

Thus, our study is confined to main-belt, non-family asteroids with semi-major axes a' in the range 2.50–3.28 AU whose proper elements are listed in the TRIAD file. 757 asteroids satisfy these requirements.

Statistical method

For each asteroid in our chosen population we calculate R , the ratio of the mean motions of Jupiter and the asteroid ($R < 1$). For each value of R we find the fractions R_2 and R_1 in the set 1/3, 3/8, 2/5, 3/7, 4/9, 5/11, 1/2, which bound R most closely from above and below. Then, using a method similar to that of Dermott¹¹, we define a parameter δ by

$$\begin{aligned} A &= (R - R_1)/(R_2 - R_1) \\ \delta &= 2(A - B) \\ B &= 0 \text{ if } A \leq 1/2 \\ B &= 1 \text{ if } A > 1/2 \end{aligned} \quad (1)$$

Thus, $-1 \leq \delta \leq +1$ and is a measure of the displacement of the asteroid from the nearest strong resonance. $\delta = 0$ corresponds to exact resonance and $|\delta| = 1$ is the greatest displacement from resonance that an asteroid can have.

Clearly the number density of asteroids systematically increases and decreases to a variable degree between the strong resonances. By defining δ we can superpose distributions to obtain a single distribution which is characteristic of the population as a whole. The relationship between δ and any other property of the asteroids or their orbits, such as their absolute magnitudes, types, diameters or proper eccentricities, can then be analysed.

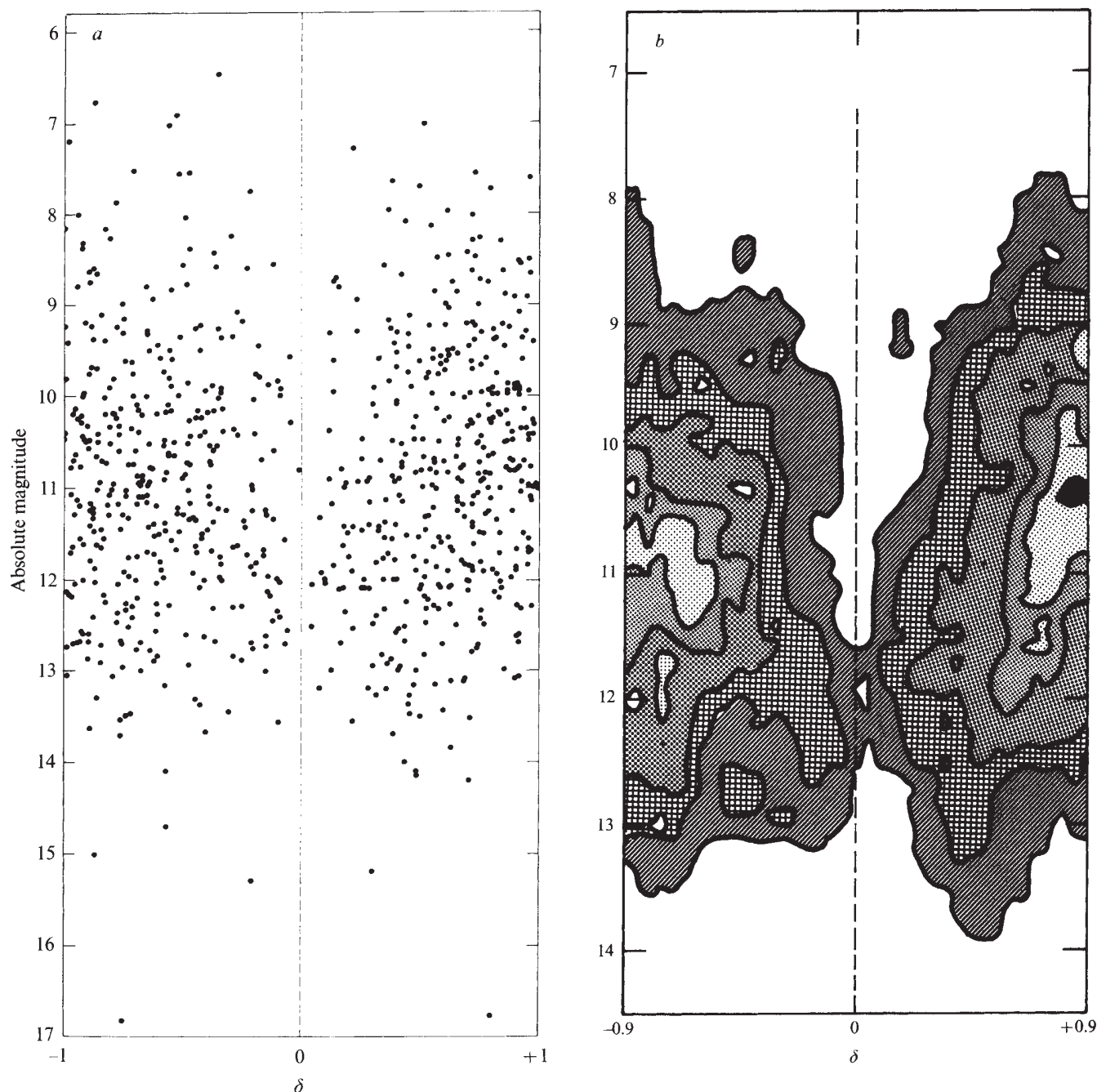


Fig. 1 *a*, A direct plot of absolute magnitude against δ (a measure of the displacement from the nearest strong resonance) for the 757 asteroids in the TRIAD file which have semi-major axes in the range 2.50–3.28 AU and which are not members of recognized families. The apparent lack of asteroids with large absolute magnitudes is an observational selection effect. *b*, A smoothed number density contour plot obtained by applying running box methods to the data shown in *a*. Note how the number density increases continuously as $|\delta|$ increases: there is no plateau. This implies that the effects of resonance pervade the whole distribution of main-belt asteroids. For absolute magnitude ≤ 10.3 the number density decreases as the absolute magnitude decreases—that is, there are relatively few asteroids of low magnitude. It is a combination of these two trends which results in the general V-shape of the contour plot.

Kirkwood gaps

Figure 1*a* plots absolute magnitude, $B(1, 0)$, against δ for all the asteroids in our chosen population. The absence of objects near $\delta = 0$, which corresponds to exact resonance, is the only accepted characteristic of Kirkwood gaps. It is useful to study a contour plot of the data shown in Fig. 1*a* and this is shown in Fig. 1*b*, which was obtained by applying a running box method to our data set. Note that as $\delta \rightarrow \pm 1$ the number density monotonically increases: the effects of resonance pervade the whole distribution of main-belt asteroids.

It has been suggested (see, for example, ref. 12) that larger objects avoid the gaps more than the smaller objects, and on first

examination Fig. 1*b* seems to confirm this. However, a combination of a decrease in number density with decreasing magnitude and an increase in number density with increasing $|\delta|$ would always give the illusion that low-magnitude objects avoid the gaps, even if the asteroids were in fact randomly distributed on either side of the resonances. Alternative methods are needed to settle this question.

By ordering our chosen population in absolute magnitude and calculating running means, we can plot the mean value of $|\delta|$ against the mean absolute magnitude (Fig. 2*a*). In Fig. 2*a*, and also Figs 2*b* and 3, the running box always contains 100 (= N) asteroids and was shifted through the chosen population with a step length of 10 asteroids.

The central limit theorem (see, for example, ref. 13) relates σ , the standard deviation of the distribution of the means of the samples, to σ_p , the standard deviation of the chosen population by

$$\sigma = \frac{\sigma_p}{\sqrt{N}} \quad (2)$$

where N is the number of objects in each sample.

Figure 2a shows that $\langle |\delta| \rangle$ increases as absolute magnitude decreases. The result is significant at $\geq 3\sigma$ level. Thus, larger objects in our chosen population do avoid the gaps more than the smaller objects. However, by ordering our chosen population by sine of proper inclination, we can also show that proper inclination increases as $\langle |\delta| \rangle$ increases (Fig. 2b). Again, the result is significant at $\geq 3\sigma$ level. Furthermore, the data shown in Fig. 2 seem to be negatively correlated, and this is confirmed in Fig. 3, which plots the mean absolute magnitude against mean sine proper inclination.

That mean absolute magnitude should decrease as mean proper inclination increases is a consequence of observational selection¹⁴. Faint asteroids have a greater probability of discovery if their inclinations are low and close to the ecliptic. Thus, although the trends shown in Fig. 2 are undoubtedly real, it is probable that only one of them is a fundamental property of

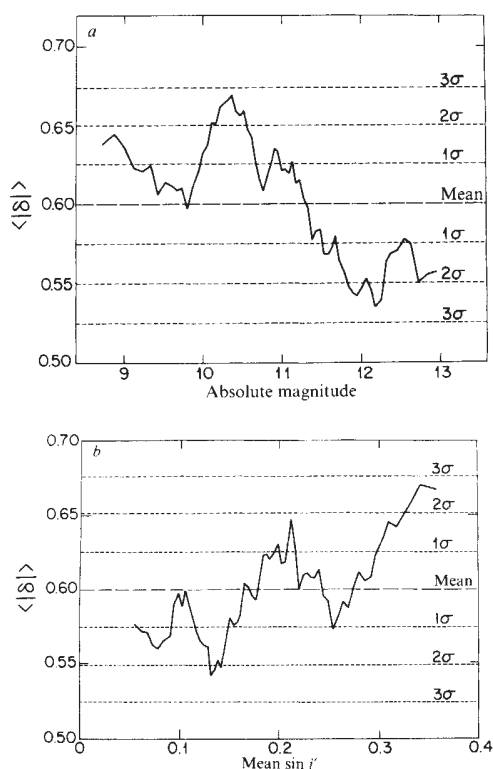


Fig. 2 A plot of the mean value of $|\delta|$ against: *a*, the mean absolute magnitude; and *b*, the mean sine proper inclination obtained by applying a running box method to the chosen population of 757 non-family asteroids which were sorted in order of increasing *a*, absolute magnitude; and *b* sine proper inclination. The σ levels are derived by the application of the central limit theorem to the distribution of the means of the samples of the population. The running box (and hence each sample) contained 100 asteroids and was stepped through the population 10 asteroids at a time. *a* Shows that low-magnitude objects do tend to have larger values of $|\delta|$, that is, that low-magnitude objects do avoid the gaps to a greater degree than do the high-magnitude objects. *b* Shows that there is a tendency for displacement from resonance to increase with inclination. In both *a* and *b* the result is significant at $\geq 3\sigma$ level.

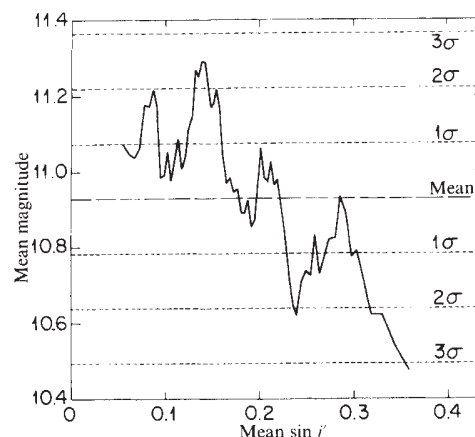


Fig. 3 A plot of the mean value of absolute magnitude against mean sine proper inclination obtained by applying a running box method to our chosen population of 757 non-family asteroids which were sorted in order of increasing sine proper inclination. The correlation is consistent with the effects of observational selection which produce an excess of high magnitude, low inclination objects¹⁴.

Kirkwood gaps and the result of resonance rather than observing procedures. We must now determine which trend is the fundamental one, and we do this by analysing a data set which is free of all observational selection effects.

Bias-free statistics

On the basis of results from the McDonald survey¹⁵, Kiang¹⁴ estimated that the catalogued asteroid population is 95% complete up to mean opposition magnitude, $B(a, 0) = 15$. Mean opposition magnitude is related to absolute magnitude by

$$B(a, 0) = B(1, 0) + 5 \log a(a - 1) \quad (3)$$

where a is the semi-major axis of the asteroid in AU. Furthermore, Kiang points out that of the 587 new objects detected in the survey only two had apparent magnitudes < 14 . To ensure complete elimination of all observational selection effects, even at the risk of being over-conservative, we adopt as an upper limit that value of absolute magnitude which corresponds to mean opposition magnitude 14 at the largest semi-major axis of our chosen population, 3.25 AU. This gives us (see equation (3)) the conditions

$$\begin{aligned} B(1, 0) &< 9.68 \\ B(a, 0) &< 14.00. \end{aligned} \quad (4)$$

Of our chosen population of 757 asteroids, 144 satisfy the above conditions. These asteroids constitute our bias-free set.

Using the bias-free set we have repeated the plot of $\langle |\delta| \rangle$ against $\langle B(1, 0) \rangle$ and have found no significant trend. The set yields no evidence that the magnitude-frequency distribution of the asteroids changes near the Kirkwood gaps. However, by repeating the plot of $\langle |\delta| \rangle$ against $\langle \sin i' \rangle$ we have found that the trend shown in Fig. 2b persists. Thus, we conclude that it is the latter which is of fundamental importance and that the trend shown in Fig. 2a is merely a consequence of observational selection effects.

We have also discovered that, in addition to the proper inclinations i' , the proper eccentricities e' also tend to increase away from the gaps. Here, these trends are best illustrated by Fig. 4 in which we plot $\langle V \rangle$ against $\langle |\delta| \rangle$, where

$$V = (e'^2 + \sin^2 i')^{1/2} \quad (5)$$

We consider it significant that the trend shown in Fig. 4 does not level off for large values of $|\delta|$, suggesting that the effects of resonance are not confined to the immediate vicinities of the resonances, but pervade the whole distribution of main-belt asteroids.

Themis family

Franklin *et al.*¹⁶ have recently suggested that the Kirkwood gaps were formed early in the history of the Solar System at a time when the asteroids moved in near-circular, near-coplanar orbits and low-velocity collisions were frequent. They consider that the present high values of inclinations and eccentricities of orbits in the asteroid belt were produced later by some process unrelated to gap formation. We can now see that their hypothesis does not account for the characteristic trends in i' and e' which we have discovered. Further evidence contrary to their hypothesis may be provided by the Hirayama families.

The 144 asteroids that constitute our bias-free set are among the largest asteroids in the main belt. Because the lifetime of an asteroid against collisions is proportional to $\sqrt{D}$, where D is the asteroid diameter¹⁷, this set also consists of some of the oldest asteroids. It would be of interest, therefore, to look for gaps in the distributions of objects which are probably relatively recent additions to the asteroid population. The larger asteroid families are generally believed^{18,19} to be the results of collisional disruption of single bodies. The Eos family is estimated²⁰ to have been formed $\geq 10^6$ yr ago, while the Themis family is thought to be somewhat older because of the larger spread of its orbital elements. A large spread in semi-major axes increases the likelihood of sufficiently strong resonances occurring within a given family. The Themis family, with 62 catalogued members⁶ lying between 3.08 and 3.18 AU, provides the best set of relatively young asteroids in which to search for gaps associated with resonances.

There are five resonances with $q \leq 11$ located within this family. When we apply a running box method to the distribution of a , we find gaps associated with the two strongest resonances (Fig. 5). In addition, the family seems to be bounded on one side by the 5/11 resonance (3.075 AU). One may question whether these gaps are real, in the sense that they represent a genuine absence of material or whether, because family members are identified by clustering of orbital elements, they represent the failure to identify asteroids in these regions as family members because their eccentricities and inclinations, but not, perhaps, their semi-major axes, have been altered by the effects of resonance. The answer to this important question may be obtained from more complete studies of background asteroids in

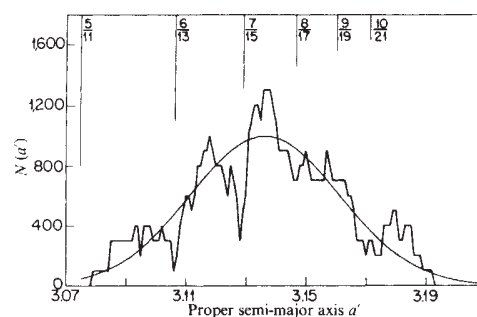


Fig. 5 The distribution of the 62 asteroids in the Themis family as a function of proper semi-major axis. The plot was obtained using a running box method with a box length of 0.01 AU and a shift of 0.001 AU between data points. $N(a')$ is the number of asteroids per unit AU contained in each box. The smooth curve represents the best gaussian fit to the data. A gaussian distribution is characteristic of catastrophic disruption²¹. The location and approximate relative strength of all resonances up to and including order 11 which lie in the region of the Themis family are shown. The family appears bounded to the left by the 5/11 resonance. The strong 1/2 resonance lies at 3.28 AU. The two strongest resonances which lie in the Themis region (6/13 and 7/15) coincide with gaps in the distribution. These gaps may be of a similar nature to the Kirkwood gaps or they could be regions where the resonances have altered the proper elements of asteroids causing them to be overlooked as potential Themis family members. Observations of the physical properties of background asteroids in these regions should help to distinguish between these hypotheses.

these gaps to determine whether or not they have the physical characteristics of Themis members.

On the basis of our studies of the Themis and other families, in particular, the Eos and Koronis families, we tentatively conclude that the gaps in their distributions of semi-major axes are real and thus that some Kirkwood gaps have been formed since the time of formation of the Solar System.

Conclusions

When the bias in the data due to observational selection is removed, there is no significant tendency for low-magnitude objects to be further away from the resonances than the high-magnitude objects. Such a magnitude effect, if found, would have supported the collisional hypothesis for the origin of the gaps.

There is a significant tendency for objects further away from the resonances to have larger eccentricities and inclinations. The process responsible for the formation of the Kirkwood gaps seems preferentially to have removed those objects near the resonances which had large values of e' and i' . Since the magnitude of the gravitational disturbing function acting on an asteroid increases with the eccentricity and the inclination of the asteroid, our results strongly suggest that it is the gravitational effect of Jupiter, acting on individual asteroids, which is responsible for the formation of the Kirkwood gaps. It is possible that such a mechanism has produced the gaps over the lifetime of the Solar System, but further studies of Hirayama families should provide more accurate measurements of the timescale of such a process.

The Kirkwood gaps are not simply regions of small asteroid number density, neither are they restricted to the immediate vicinities of the resonances, for the effects of resonance seem to pervade the entire belt of asteroids between the 1/3 resonance and Jupiter itself. The characteristics we have described should place severe constraints on any theory of the formation of the Kirkwood gaps. They also challenge our understanding of the dynamical stability of the Solar System.

We thank P. Goldreich, J. Gradie and T. C. Van Flandern for helpful discussions and B. Zellner for providing us with a copy of the TRIAD file. This research was supported by NSF grant AST 79164-74-A01.

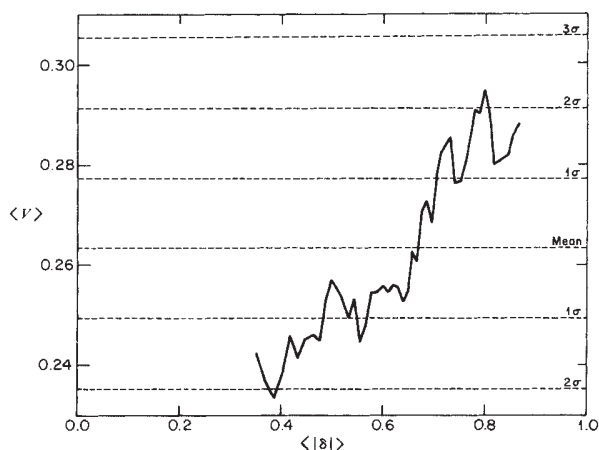


Fig. 4 A plot of the mean value of $V = (e'^2 + \sin^2 i')^{1/2}$ against the mean value of $|\delta|$ obtained by applying a running box method to the bias-free set of 144 non-family asteroids which were sorted in order of increasing $|\delta|$. For this plot, the running box contains 50 asteroids and there is a shift of two asteroids between data points. Note that the plot has no plateau: the trend is present throughout the distribution. This tendency for both e' and i' to increase with increasing $|\delta|$ is a fundamental property of the Kirkwood gaps and is indicative of the effects of gravitational forces acting on individual asteroids as a possible mechanism for the origin of the Kirkwood gaps.

Received 22 December 1980; accepted 5 February 1981.

- Kirkwood, D. *Meteorite Astronomy* (Lippincott, Philadelphia, 1867).
- Greenberg, R. & Scholl, H. in *Asteroids* (ed. Gehrels, T.) 310–333 (University of Arizona Press, 1979).
- Message, P. J. in *The Theory of Orbits in the Solar System and in Stellar Systems* (ed. Contopoulos, G.) 197–222 (Academic, London, 1966).
- Schweizer, F. *Astr. J.* **74**, 779–788 (1969).
- Wiesel, W. E. *Celestial Mech.* **13**, 3–37 (1976).
- Sinclair, A. T. *Mon. Not. R. astr. Soc.* **142**, 289–292 (1969).
- Brouwer, D. & Clemence, G. M. *Methods of Celestial Mechanics* (Academic, New York, 1961).
- Williams, J. G. thesis, Univ. California (1969).
- Williams, J. G. in *Asteroids* (ed. Gehrels, T.) 1040–1063 (University of Arizona Press, 1979).
- Hirayama, K. *Proc. phys.-math. Soc. Japan Ser. II*, **9**, 354–361 (1918).
- Dermott, S. F. *Nature phys. Sci.* **244**, 18–21 (1973).
- Zellner, B. & Howell, E. in *Comets, Asteroids, Meteorites; Interrelations, Evolution and Origins* (ed. Delsemme, A. H.) 185–197 (University of Toledo, 1977).
- Lindgren, B. W. *Statistical Theory* (Macmillan, New York, 1960).
- Kiang, T. *Icarus* **5**, 437–449 (1966).
- Kuiper, G. P. *et al. Astrophys. J. Suppl.* **3**, 289–428 (1958).
- Franklin, F. A., Lecar, M., Lin, D. N. C. & Papaloizou, J. *Icarus* **42**, 271–280 (1980).
- Chapman, C. R. in *Asteroids* (ed. Gehrels, T.) 25–60 (University of Arizona Press, 1979).
- Gradie, J. & Zellner, B. *Science* **197**, 254–255 (1977).
- Gradie, J. thesis, Univ. Arizona (1978).
- Steins, K. A. *Astr. Zh.* **33**, 246–250 (1956).
- Wiesel, W. *Icarus* **34**, 99–116 (1978).

Interacting genes in nematode dauer larva formation

Donald L. Riddle, Margaret M. Swanson & Patrice S. Albert

Division of Biological Sciences, University of Missouri, Columbia, Missouri 65211, USA

The dauer larva of Caenorhabditis elegans is a developmentally arrested stage induced by starvation or overcrowding. Mutant genes controlling the ability to form dauer larvae interact in a way which allows them to be ordered in a pathway. Mutant phenotypes suggest that the pathway corresponds to neural processing of environmental stimuli.

To the extent that the development of organisms is controlled by genes which act in sequence¹, it should be possible to determine a hierarchy of genes affecting the expression of certain traits. What follows is an account of a study, in the language of classical genetics, of the hierarchical relationship between the genes involved in the formation of the so-called dauer larval stage of the free-living soil nematode *Caenorhabditis elegans*. The genetic and anatomical simplicity of the organism^{2–5} fit it for such a study. Our work shows, however, that the expected hierarchy is surprisingly intricate.

Larvae of *C. elegans* undergo four moults (Fig. 1), at which the cuticle is replaced, but at the second may enter the 'dauer' or 'persistent' (German: *dauer* = hard, enduring) state⁶ in response to starvation or overcrowding. Thus arrested in development but specialized for survival and dispersal⁷, dauer larvae survive exposure to detergents and other chemicals⁸, do not feed because their mouths are plugged (Fig. 2) and may survive for months until they encounter food, moult and resume development. The duration of the dauer stage seems not to affect post-dauer longevity, suggesting that ageing is interrupted⁹.

Table 1 Suppression of dauer-constitutive mutations by dauer-defectives

Dauer defective mutants	Multiply-marked dauer-constitutive strains						
	DR194 (<i>daf-11 V</i>)	DR189 (<i>daf-8 I</i>)	DR133 (<i>daf-7 III</i>)	DR202 (<i>daf-14 IV</i>)	DR204 (<i>daf-1 IV</i>)	CB2196 (<i>daf-4 III</i>)	DR130 (<i>daf-2 III</i>)
CB1376 (<i>daf-3 X</i>)	+	+	+	+	+	+	–
CB1386 (<i>daf-5 II</i>)	+	+	+	+	+	+	–
DR20 (<i>daf-12 X</i>)	+	+	+	+	+	+	–
DR25 (<i>daf-20 X</i>)	+	+	+	+	+	–	+
DR26 (<i>daf-16 I</i>)	+	+(60%)	+(20%)	+(25%)	+(35%)	–	+
CB1377 (<i>daf-6 X</i>)	+(70%)	+(30%)	+	+(35%)	–	–	–
DR27 (<i>daf-1 I</i>)	+	+	+	–	–	–	+
CB1375 (<i>daf-18 IV</i>)	+	+	–	–	–	–	–

+ Indicates that, in a particular constitutive-defective double mutant, the constitutive phenotype is suppressed (the defective mutation is epistatic). – Indicates gene pairs in which the constitutive is epistatic (no suppression of the constitutive phenotype). In cases where suppression is not complete (*daf-6* and *daf-16*), the fraction of the double-mutant population which escapes dauer formation is given in parentheses as per cent suppression. As the only available mutant alleles of these genes are slightly leaky dauer-defectives, a low residual level of gene activity may account for reduced efficiency of suppression. Dauer-defective strains are all single mutants with allele numbers corresponding to the strain numbers given. Gene names and linkage groups are given in parentheses. The multiply-marked dauer-constitutive strains have the following genotypes, where allele numbers are given in parentheses after the corresponding gene name: DR194, *dpy-11(e224)daf-11(m47)unc-39(e257)V*; DR189, *dpy-5(e61)daf-8(e1393)unc-29(e1072)II*; DR133, *dpy-1(e1)daf-7(e1372)unc-32(e189)III*; DR202, *unc-24(e138)daf-14(m77)IV*; DR204, *daf-1(m40)dpy-4(e1166)IV*; CB2196, *daf-4(e1364)unc-32(e189)III*; DR130, *dpy-1(e1)daf-2(e1370)unc-32(e189)III*. Because it was impossible to predict in advance if a double mutant would be constitutive or defective (or even viable), closely linked visible markers were used to identify the presence or absence of *daf* alleles in newly constructed strains. To construct the strains, males homozygous for a dauer-defective (suppressor) mutation were crossed with *ts* dauer-constitutive hermaphrodites carrying closely linked, flanking mutations. Heterozygous F₁ hermaphrodites produced in these crosses were wild type in phenotype, and segregated dumpy, uncoordinated animals by 'selfing' at 15 °C. The dauer-constitutive phenotype is not expressed at 15 °C. In each test, 15 marked segregants were incubated at 25 °C to determine if suppression occurred. If the suppressor gene is unlinked, three-quarters of the marked animals are expected to carry the suppressor gene in either the heterozygous or homozygous state. All such animals segregate homozygous progeny which, in the case of suppression, will grow to maturity at 25 °C. If there is no suppression, all progeny become dauer larvae. In cases where closely linked genes were tested, recombinants carrying both mutations were selected by using an intervening visible marker as in the case of *daf-8* and *daf-16*. Starting with a heterozygote (++ *daf-16/daf-8 unc-29* +), a *daf-8 daf-16* recombinant was selected by picking F₁ progeny which are dauer-constitutive, but not uncoordinated (*daf-8 + daf-16/daf-8 unc-29* +). At 15 °C, this heterozygote segregated the double homozygote required to test the epistatic relationship of *daf-8* and *daf-16* at 25 °C. Crosses with the *daf-10* mutant are not included here. Appropriate double mutants were constructed by another method¹², because homozygous *daf-10* males do not mate¹¹. All the constitutive mutations except *daf-4* have been combined with *daf-10*, and each failed to be suppressed.

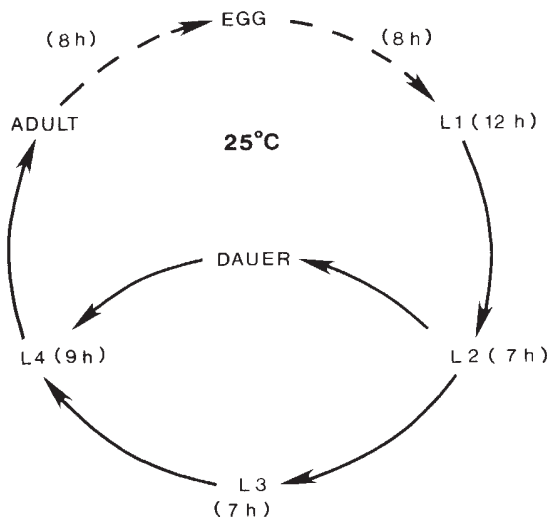


Fig. 1 The life cycle of *C. elegans*. The duration of each developmental stage during growth at 25°C on Petri plates seeded with *Escherichia coli* strain OP50 (ref. 2) is given in parentheses. The solid arrows represent moults. Moulting times were determined by monitoring pharyngeal pumping in synchronized populations¹². Wild-type dauer larvae obtained from starved populations initiate pharyngeal pumping 2–3 h after being placed in food. The moult to the L4 stage occurs after an additional 8 h. The 8-h time from egg to L1 is the interval between egg laying and hatching. The entire life cycle requires 3 days at 25°C.

Such stages of arrested development occur in other nematodes, and the similarity with the obligate (usually infective) form of many pathogenic parasites should not be entirely overlooked¹⁰.

In our genetic studies, we have isolated mutants of *C. elegans* of two kinds—'defective' mutants which do not produce dauer larvae even when starved¹¹, and 'constitutive' mutants which produce dauer larvae even with abundant food¹². The genes concerned are named *daf* (for 'dauer larva formation').

We have screened clones for defective mutants by treating nematodes with the mutagen ethyl methane sulphonate (EMS) and selecting clones of F₂ hermaphrodite progeny which do not form dauer larvae after starvation; a total of at least 15 genes is involved^{11,13}. We have obtained 38 constitutive mutants from nematodes treated with either EMS or DES (diethyl sulphate) by selection for dauer larvae with 1% SDS¹⁴.

Thirty-three of the constitutive mutants are *ts* (temperature-sensitive), while five are 'leaky' (incompletely penetrant) at all growth temperatures. The best of the *ts* mutants grow normally at 15°C, but all larvae grown at 25°C become dauers at the second moult. The dauer larvae cannot recover until the incubation temperature is shifted down to 15°C. The *ts* mutants are all recessive, and they define seven complementation groups (see Table 1). SDS selection can only yield conditional or leaky mutants which are able to recover and reproduce. The isolation of unconditional (non-recovering) dauer-constitutive mutants will be described elsewhere.

Epistatic ordering of genes

A relationship between constitutive and defective mutations was detected by selecting revertants of *ts* dauer-constitutives, *daf-2 daf-4* and *daf-7* (Table 2). Homozygous revertants, arising in the F₂ generation after mutagenesis, grow and reproduce at 25°C, while parental-type larvae become dauers and stop growing. Most revertants not only escape dauer formation, but they fail to form dauers even when starved. Crossing the revertants with the parental strain revealed that they were double mutants which carried not only the parental constitutive mutation, but also were homozygous for an epistatic, recessive dauer-defective mutation. Thus, some dauer-defective mutations can act as suppressors of dauer-constitutives.

Twenty-five independently isolated revertants were distributed among six complementation groups (Table 2) and some

revertants were alleles of dauer-defective genes which had already been genetically mapped (Fig. 3). To determine which dauer-constitutive mutants could be suppressed by various defectives, we constructed a series of multiple mutants by genetic crosses (Table 1). One interpretation of these data could be that constitutive and defective mutations produce competitive influences on dauer larva formation in a double mutant. The more extreme alleles might be epistatic to less extreme alleles, so that the cross-suppression tests rank the mutations quantitatively in order of severity. Several observations argue against this hypothesis as a general explanation for the observed interactions. First, there is no uniform gradient of suppression. For example, the *daf-12* mutation suppresses *daf-4* but not *daf-2* while the *daf-20* mutation exhibits the opposite pattern. Second, ranking the *ts* mutants by the degree of expression at intermediate temperatures does not produce a correspondence to the suppression pattern¹². Based on the degree of expression of the *ts* constitutive strains at a range of temperatures between 15 and 25°C, the *daf-7* strain was judged to carry the most severe *ts* allele, while the *daf-14* strain carries the least severe allele. Third, *daf-16* provides an absolute block to the most

Table 2 Origin of dauer-defective mutations

Gene and linkage group	Reference allele	No. of alleles collected*			Estimated mutation frequency
		By reversion†	Other‡		
<i>daf-10 IV</i>	<i>e1387</i>	—	—	2	5 × 10 ⁻⁴
<i>daf-18 IV</i>	<i>e1375</i>	—	—	1	
<i>daf-17 I</i>	<i>m27</i>	—	5	—	1 × 10 ⁻⁴
<i>daf-6 X</i>	<i>e1377</i>	—	—	1	
<i>daf-16 I</i>	<i>m26</i>	—	1	—	2 × 10 ⁻⁵
<i>daf-20 X</i>	<i>m25</i>	—	2	—	5 × 10 ⁻⁵
<i>daf-3 X</i>	<i>e1376</i>	5	2	1	3 × 10 ⁻⁴
<i>daf-5 II</i>	<i>e1386</i>	7	2	2	5 × 10 ⁻⁴
<i>daf-12 X</i>	<i>m20</i>	1	—	—	5 × 10 ⁻⁵

* Complementation tests were performed by crossing homozygous or hemizygous dauer-defective males with dumpy (*dpy*) dauer-defective (*daf*) hermaphrodites. Starved populations derived from cloned, non-dumpy F₁ progeny were scored for the presence or absence of dauer larvae both visually and by SDS treatment^{11,13}. Populations arising from heteroallelic F₁ progeny do not form dauers. Complementation *daf* mutations, on the other hand, allow many dauers to be formed even though some homozygous dauer-defective progeny are segregated in the F₂ generation. Because homozygous *daf-10* males do not mate¹¹, heterozygous *daf-10/+* males were crossed with *dpy daf* hermaphrodites. In these crosses, half the F₁ progeny produce dauer-defective populations when the *daf* mutations are allelic.

† To obtain revertants of *ts* dauer-constitutive mutations, strains CB1372, *daf-7(e1372)III*; CB1364, *daf-4(e1364)III*; CB1370, *daf-2(e1370)III*; CB1368, *daf-2(e1368)III* and CB1391, *daf-2(e1391)III* were mutagenized with EMS (ref. 6). In each experiment 4 adult hermaphrodites were placed on each of 50 100-mm diameter Petri plates seeded with OP50 and allowed to produce about 200 F₁ progeny per plate. After an initial 6 days of incubation at 15°C, adult F₁ progeny were shifted to 25°C for an additional 4 days. Adult F₂ revertants were cloned and re-tested for growth at 25°C, and for the dauer-defective phenotype after starvation. To assure independence of the mutations, no more than one revertant clone from each plate was saved. Homozygous revertant males were backcrossed to *dpy daf* hermaphrodites. The appearance of dauer-constitutive cross progeny in the F₁ demonstrates the continued presence of the parental dauer-constitutive allele in the revertant strain, as well as the recessive character of the suppressing dauer-defective mutation. The dauer-constitutive alleles were then removed from the genetic background of each revertant by crossing the homozygous revertant males with strains carrying *dpy* markers closely linked to the constitutive locus. Doubly homozygous *dpy;daf* F₂ segregants which no longer carried the constitutive mutation were saved for complementation tests. The *dpy;daf* strains were again crossed with wild-type males to remove the *dpy* marker. In each case homozygous or hemizygous males derived from a non-dumpy, dauer-defective segregant were used for cross-suppression tests (see Table 1).

‡ Alleles *e1387*, *e1376* and *e1386* were obtained by cloning F₂ progeny of mutagenized N2 hermaphrodites as described elsewhere¹¹. All mutants were subsequently backcrossed and re-segregated before complementation and genetic mapping. The *daf-18(e1375)III* mutation was discovered in strain CB1062, *vab-3 X*, and removed from the *vab-3* background by crossing with wild-type males and re-segregating the unlinked *e1375* allele. The *e1377* mutant was isolated after EMS mutagenesis by Dr James A. Lewis. The frequency of mutation in each gene in standard conditions of EMS mutagenesis² was estimated by dividing the total number of mutants obtained (columns 3–6) by twice the number of F₁ broods examined.

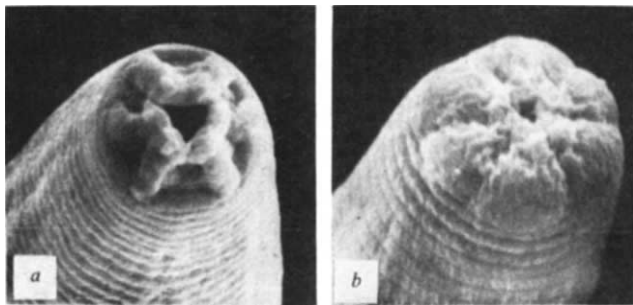


Fig. 2 External morphological differences between an L2 larva (a) and a dauer larva (b) are shown by scanning electron micrographs of the nose. The field width in each micrograph is $7.2 \mu\text{m}$. An internal plug which completely closes the mouth of the dauer (b) is revealed in transmission electron micrographs of transverse sections (not shown). The dauer larva is thinner than the L2, and has an increased specific gravity due to radial shrinkage of the body at the dauer-specific moult⁷. In preparation for scanning electron microscopy, larvae were fixed overnight in cold 3% glutaraldehyde in 0.1 M phosphate buffer, pH 7.1, and post-fixed for 6 h in cold 1% buffered OsO_4 . Specimens were dehydrated in a graded series of ethanol, critical point dried in CO_2 , mounted on copper adhesive tape and gold coated. Photographs were taken at 10 kV on a Jeol JSM-1 scanning electron microscope.

extreme (non-recovering) constitutive mutations, although it does not suppress *ts* alleles of certain constitutive genes (in preparation).

A functional sequence

The data suggest that there is an order to the functions controlled by the *daf* genes, and that the constitutive and defective mutations affect a common sequence. On the basis of the suppression patterns (Table 1) the pathway in Fig. 4 can be constructed with individual genes assigned to the control of sequential steps. If the constitutive mutants were simply blocked in entry to the L3 stage (see Fig. 1), while defective mutants were blocked in a separate sequence to the dauer larva, the constitutive-defective double mutants could not continue larval development at all. Instead, the data indicate that the dauer-defective mutants are blocked in the pathway to the dauer stage while the dauer-constitutives generate a false, internal signal which causes the mutants to form dauer larvae even in the absence of appropriate environmental cues. If the pathway is blocked after the false signal, a double mutant will be dauer-defective. However, if the false signal is generated after the block, a double mutant will be dauer-constitutive.

An unambiguous order of function can be assigned to most dauer-constitutive genes. However, *daf-2* seems to be in a distinct branch of the pathway. The *daf-16* and *daf-20* mutations suppress *daf-2* but not *daf-4*, while *daf-3*, *daf-5* and *daf-12* suppress all constitutives except *daf-2*. A block in the upper branch of the pathway is both necessary and sufficient to produce a dauer-defective phenotype. Therefore, we conclude that the lower pathway specified by *daf-2* does not function in these conditions. This second pathway may represent the response to different environmental signals, or it may function only in *daf-2* mutants.

Our goal is to establish specific behavioural or morphological correlates with the mutant steps in the pathway. Two steps, represented by *daf-10* and *daf-6*, are correlated with chemotaxis defects, and both mutants exhibit ultrastructural alterations¹¹ in specific neurones which have previously been implicated in the chemosensory response to salts^{15,16}. These dauer-defective mutants exhibit pleiotropic defects, both in chemosensory behaviour and in neuroanatomy. Nevertheless, ultrastructural evidence suggests that one or more amphidial neurones mediate a sensory signal for dauer larva formation¹¹. The amphids are a pair of sense organs consisting of 12 neurones, 8 of which extend afferent processes anteriorly to openings on the lateral sides of the mouth^{17,18}.

The genetic pathway may correspond to a neural pathway involved in detection of an environmental signal by sensory

receptors followed by transduction of the signal to neurosecretory cells. According to this hypothesis, neurosecretion would then alter the hormonal balance in the nematode to initiate dauer larva morphogenesis. The final step in the pathway may represent defects in the reception of the neurosecretory signal. Underlying the neural pathway would be the developmental processes which produce a functional set of nerve cells. For example, a mutation affecting morphogenesis of sensory neurones would affect the pathway at a position corresponding to that sensory function. Thus, we expect that many of the genes exert their primary effects on development well before the pathway is actually set in motion physiologically. This contention is supported by the results of temperature-shift and temperature-pulse experiments which define temperature-sensitive periods before the L2 stage for six of the *ts* dauer-constitutive mutants¹².

One or more steps in the pathway may involve amphid function. The *daf-6* mutant exhibits amphid defects as does *daf-10*, but other neurones in the *daf-6* mutant are also affected¹¹. None of the other mutations positioned in the pathway exhibits the behavioural abnormalities characteristic of the non-chemotactic dauer-defective mutants, *daf-6* and *daf-10*.

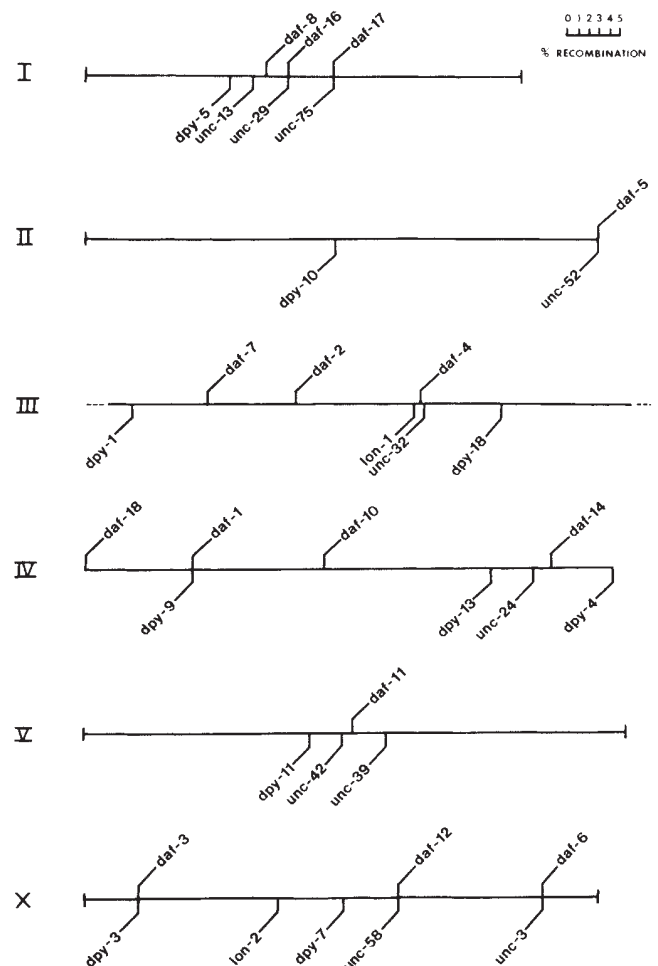
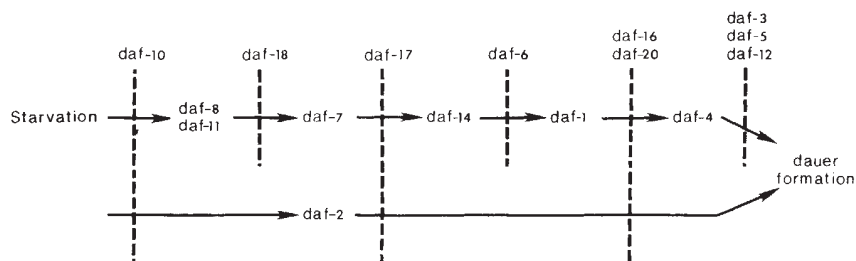


Fig. 3 Simplified genetic map of *C. elegans* showing the six linkage groups with *daf* loci labelled above the lines. Other markers used in mapping crosses and in cross-suppression tests (Table 1) are shown below the lines. Linkage relationships are as described previously²¹. Gene designations are *daf* (dauer larva formation), *dpy* (dumpy), *lon* (long) and *unc* (uncoordinated). The locations are shown only for mutations which have been unambiguously ordered. When two genes are joined at the same location, no recombination has been detected between them. The *daf-20* gene is X-linked, but is not shown because its location has not been determined.

Fig. 4 A simple genetic pathway for dauer larva formation, one of several possible representations, based on epistatic relationships between dauer-constitutive mutations and dauer-defective mutations. Gene names of dauer-constitutives are given in the pathway itself to represent points where false signals may be initiated. Mutations in dauer-defective genes block the pathway at the positions shown by dashed lines. The *daf-10* mutation is drawn to block both branches of the pathway on the basis of indirect evidence. First, both *daf-8* and *daf-2* are epistatic to *daf-10*, and second, the *daf-10* mutation blocks recovery of both *daf-2* and *daf-8* dauer larvae (see text).



However, if the gene order corresponds to a pathway of neurotransmission, electron microscopic analysis of *daf* mutants positioned near the end of the pathway may reveal anatomical defects in neurosecretory cells.

Genetic complexity of the pathway

C. elegans possesses three to four thousand essential genes^{2,19}. We estimate, based on mutation frequency and the fraction of new mutations which recur in previously characterized genes (in preparation), that there are ~100 genes affecting dauer larva formation. Dauer-constitutive mutants are rare, occurring at an aggregate frequency of 0.2% or about four times the mutation frequency per average gene in standard conditions of EMS mutagenesis^{2,13}. All seven of the constitutive genes detected by our mutant selection are represented by more than one allele. Therefore, we conclude that we are approaching saturation of the *ts* dauer-constitutive mutant class.

Two other experiments support this view. First, we have used dauer-defective mutants late in the pathway (*daf-16* and *daf-20*) as parent strains for selecting a restricted class of dauer-constitutive mutants which might define even later steps in the pathway (see Fig. 4). After mutagenesis with EMS or DES, the progeny of nearly 9,000 F₁ animals were screened. This population is of sufficient size to detect four or five alleles of the average gene in *C. elegans*. In fact, five new alleles of *daf-4* were found, but no new dauer-constitutive genes were identified. Second, mutants with maternal effects have not been overlooked. All eight alleles of *daf-1* share this property. Homozygous *daf-1* animals do not form dauer larvae if their maternal parent was heterozygous (+/*daf-1*). Selection for maternal-effect mutants (not expressed until the F₃ after mutagenesis) produced six new alleles of *daf-1*, but no other constitutive genes with a maternal effect.

Although the genetic evidence suggests that there may not be many more than 7 genes of the *ts* constitutive class, we estimate that there may be as many as 100 genes of the defective type¹³. It is likely that more defective genes will be positioned within the pathway, since four of the nine genes in the pathway are represented by only a single isolate (see Table 2) and other dauer-defective mutants have not been tested for their suppression properties. The frequency of dauer-defective mutants obtained by EMS mutagenesis of wild-type *C. elegans* is 5%. Many of these mutations may exert their effects indirectly as in the case of *daf-10* and *daf-6*, two mutations which affect neural development.

Other genes do not affect the signal, but do affect the subsequent process of dauer larva morphogenesis. For example,

a *daf-13* mutant forms apparently normal dauer larvae in response to starvation, but the dauers are sensitive to SDS treatment, possibly because of a defect in cuticle maturation (unpublished results). When the *daf-13* mutation is combined with the dauer-constitutive, *daf-2*, the constitutive phenotype is unaltered except that the constitutively formed dauer larvae are sensitive to SDS. We believe *daf-13* to be representative of a class of 'process' mutations which do not affect the signal pathway and do not exhibit epistatic interactions with dauer-constitutive mutations.

There are few, if any, genes which are involved specifically in exit from the dauer stage. Selection (by resistance to SDS) for mutants which are abnormally slow in dauer larva recovery produced no new mutant types. The only mutants found were dauer-constitutives (which recover slowly) and leaky dauer-defectives (affected in both entry and exit). This suggests that the genetic requirements for recovery from the dauer stage are similar to the requirements for entry. To determine if all dauer-defective genes specifying the entry pathway also affect exit from the dauer stage, we constructed the following three strains in which the constitutive gene, *daf-1*, is epistatic: *daf-1 daf-10*, *daf-1; daf-17* and *daf-1; daf-6*. In both the *daf-10* and *daf-6* double mutants, recovery of constitutively formed dauers (on temperature-shift to 15°C) is almost completely inhibited, showing that function of these genes is required for both entry and exit, but in the case of the *daf-1; daf-17* double mutant, dauer larva recovery is at most only marginally inhibited. Thus, the genes required for exit from the dauer stage may be simply a subset of the genes required for entry.

Dauer larva formation seems to be relatively unaffected by most of the animal's genes, allowing us to study this sequence in relative isolation from the rest of development. The properties of this system permit an unusual application of classical genetic methods to the analysis of a developmental pathway. Elucidation of the physiology of dauer larva formation will require further work to identify the specific molecules which act as environmental cues, the neurones involved in signal processing, what neurotransmitters these cells use and what hormones mediate the response. As a formal genetic problem, dauer larva development seems about as complex as that of bacteriophage T4 (ref. 20), and it is therefore a reasonable goal to characterize its genetic specification completely.

This work was supported by USPH grant HD11239 and National Research Service Award HD05654. We thank Susan Brown for the scanning electron micrographs, Sydney Brenner, Robert Waterston and Gunter Stent for helpful discussion, and Betty Hearn for manuscript preparation.

Received 8 October 1980; accepted 13 January 1981.

1. Beadle, G. W. & Ephrussi, B. *Genetics* **22**, 76–86 (1937).
2. Brenner, S. *Genetics* **77**, 71–94 (1974).
3. von Ehrenstein, G. & Schierenberg, E. in *Nematodes as Biological Models* Vol 1 (ed. Zuckerman, B. M.) 2–71 (Academic, New York, 1980).
4. White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. *Phil. Trans. R. Soc. B* **275**, 327–348 (1976).
5. Sulston, J. E., Albertson, D. G. & Thomson, J. N. *Dev. Biol.* **78**, 542–576 (1980).
6. Bovien, P. *Vidensk. Meddr dansk. naturh. Foren.* **101**, 1–113 (1937).
7. Cassada, R. C. & Russell, R. L. *Dev. Biol.* **46**, 326–342 (1975).
8. Cassada, R. C. & Russell, R. L. *Fedn Proc.* **33**, 1476 (1974).
9. Klass, M. & Hirsh, D. *Nature* **260**, 523–525 (1976).
10. Evans, A. A. F. & Perry, R. N. in *The Organization of Nematodes* (ed. Croll, N. A.) 383–424 (Academic, New York, 1976).

11. Albert, P. S., Brown, S. & Riddle, D. L. *J. comp. Neurol.* (in the press).
12. Swanson, M. M. & Riddle, D. L. *Dev. Biol.* (in the press).
13. Riddle, D. L. *Stadler Genet. Symp.* **9**, 101–120 (1977).
14. Cassada, R. C. in *Developmental Biology: Pattern Formation and Gene Regulation* (eds McMahon, D. M. & Fox, C. F.) 539–547 (Benjamin, Menlo Park, 1975).
15. Lewis, J. A. & Hodgkin, J. A. *J. comp. Neurol.* **172**, 489–510 (1977).
16. Ward, S. *Proc. natn. Acad. Sci. U.S.A.* **70**, 817–821 (1973).
17. Ward, S., Thomson, N., White, J. G. & Brenner, S. *J. comp. Neurol.* **160**, 313–338 (1975).
18. Ware, R. W., Clark, D., Crossland, K. & Russell, R. L. *J. comp. Neurol.* **162**, 71–110 (1975).
19. Herman, R. K. & Horvitz, H. R. in *Nematodes as Biological Models* Vol 1 (ed. Zuckerman, B. M.) 228–261 (Academic, New York, 1980).
20. Wood, W. B., Edgar, R. S., King, J., Lielausis, I. & Henninger, M. *Fedn Proc.* **27**, 1160–1166 (1968).
21. Herman, R. K., Horvitz, H. R. & Riddle, D. L. *Genet. Maps* **1**, 183–193 (1980).

Left-handed DNA in restriction fragments and a recombinant plasmid

Jan Klysik, Steven M. Stirdivant, Jacquelyn E. Larson, Phillip A. Hart & Robert D. Wells

University of Wisconsin, Department of Biochemistry, College of Agricultural and Life Sciences, Madison, Wisconsin 53706, USA
and School of Pharmacy, University of Wisconsin, Madison, Wisconsin 53706, USA

Circular dichroism and ^{31}P -NMR on synthetic oligomers of (dC-dG) inserted within DNA restriction fragments indicate that the right-handed B-structure can exist in close proximity to the left-handed Z-structure. Also, this salt-induced transition to Z-form in a small (dC-dG) segment (1.3%) of a recombinant plasmid markedly influenced the supercoil of the plasmid. These observations have implications for the postulated role of naturally occurring related simple sequences in the regulation of gene activity.

DUPLEX DNA polymers and oligomers containing a strictly repeating dG and dC sequence have been known for some time to display unusual conformational properties¹. High molecular weight (dG-dC)_n · (dG-dC)_n has been synthesized² and characterized by a variety of biochemical and physical approaches (reviewed in refs 1, 3). The presence of high concentrations of salt (NaCl, NaClO₄, MgCl₂) or ethanol causes this polymer, but not related synthetic polymers, to undergo a marked structural transition as revealed by circular dichroism (CD) and Raman spectroscopy⁴⁻⁷. This cooperative transition has more recently been interpreted (see below) as a conversion of the right-handed helix into a left-handed structure, an observation foreshadowed by the demonstration that (dI-dC)_n · (dI-dC)_n had a very unusual left-handed conformation at normal salt concentrations and could undergo a transition to a typical right-handed DNA-B conformation at high humidities⁸⁻¹⁰.

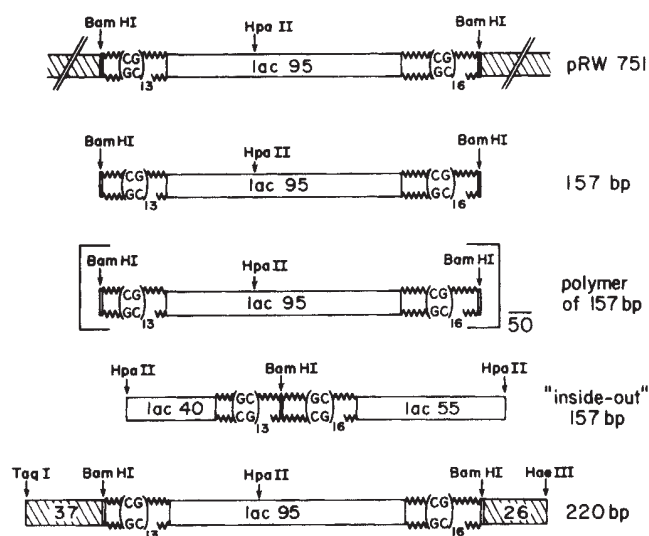
Recent X-ray crystallography studies on short oligomers [d(CpGpCpG) and d(CpGpCpGpCpG)] revealed the very unusual Z- and Z'-DNA conformations¹¹⁻¹⁵. The Z-conformation differs from the typical right-handed DNA-B structure in that the helix is left-handed with 12 base pairs per turn, the sugar conformations are not all the same but dG is C3'-endo and

dC is C2'-endo and the G residues are in a *syn*, not the orthodox *anti*, conformation. The Z-conformation may correspond to the high salt form of (dG-dC)_n · (dG-dC)_n with the 'inverted' CD spectrum¹¹, particularly as Raman spectroscopy of the crystals used for the X-ray analyses demonstrates the correlation with the high salt form of (dG-dC)_n (A. Rich, personal communication). Also, recent ^{31}P -NMR studies (refs 16, 17 and J. S. Cohen, personal communication) on oligomers with strictly alternating dG and dC residues demonstrates the salt-induced conformational transition from a typical phosphodiester structure (in low salt) to an alternating structure (in high salt). The alternating backbone conformation is presumed to be of the Z-type. It may be that, as proposed^{11,18}, any DNA with a perfectly alternating sequence of purines and pyrimidines adopts a Z-type conformation.

Z-type DNA in a natural genome

The relevance of these observations to natural DNA and their possible implications for the regulation of gene activity is still largely speculative. We therefore decided to construct a recombinant DNA in which poly(dG-dC) was cloned in restriction fragments of plasmid DNA. This approach, based on our

Fig. 1 DNAs used in these studies. pRW751 is a recombinant plasmid derived from pBR322 (ref. 29) which contains a 157-base pair (bp) insert into the 'filled in' *Bam*HI site. The cloning procedure, preparation of plasmid DNA and preparation and characterization of DNA restriction fragments were generally performed as previously described^{1,21,22,24,31}. The 157-base pair fragment was sequenced essentially as described elsewhere²³; the *Bam*HI ends were labelled with ^{32}P and the fragment was cleaved with *Hpa*II. The DNA sequence revealed the presence of the perfectly alternating (dC-dG) sequence shown in this figure along with the presence of the *lac* sequence. The localization of the 26- and 32-base pair (dC-dG) arms relative to the *lac* 95-base pair fragment is precisely known. However, the orientation of the entire 157-base pair insert with respect to the vector is not known. The polymer of the 157-base pair fragment was prepared by a large-scale T4 DNA ligase reaction with the 157-base pair monomer in conditions described previously²² using 100 µg of the 157-base pair fragment. The number average molecular weight (50-mer) of the polymers which resulted from this reaction was established by gel electrophoresis and, more rigorously, by spectrophotometric analysis of the RPC-5 chromatography used in the preparation of the 'inside-out' 157-base pair fragment. This fragment was prepared by *Hpa*II cleavage of the polymer of the 157-base pair fragment. The 'inside-out' molecules are, in fact, a mixture consisting of fragments which are 138, 157 and 176 base pairs long. This mixture was purified from the 'half molecules' (70 and 87 base pairs in length) which were derived from the termini of the polymer. Gel electrophoretic analysis of the purified 'inside-out' 157-base pair molecules showed no detectable 'half molecules' (<1%). The 220-base pair fragment was isolated from pRW751 by a three-step procedure; first, 6 mg of pRW751 were digested with *Hind*II + *Hind*III and the resulting products were separated on a preparative polyacrylamide gel. The 780-base pair fragment was extracted from the gel and cleaved with *Taq*I + *Hae*III and the products of this digestion were then fractionated on a preparative polyacrylamide gel. The 220-base pair fragment was extracted from the gel and further purified by column chromatography on DEAE-Sephacel (0.5 × 0.5 cm) in 10 mM Tris-HCl (pH 7.6), 0.1 mM EDTA. The column was washed with 3 ml of the same buffer and the 220-base pair fragment was eluted with 2.0 M NaCl in 10 mM Tris-HCl (pH 7.6) and 0.1 mM EDTA. The orientation of the 37- and 26-base pair segments of pBR322 on the termini of this molecule relative to the 157-base pair segment was not determined. pBR322 segments are designated with cross-hatching.



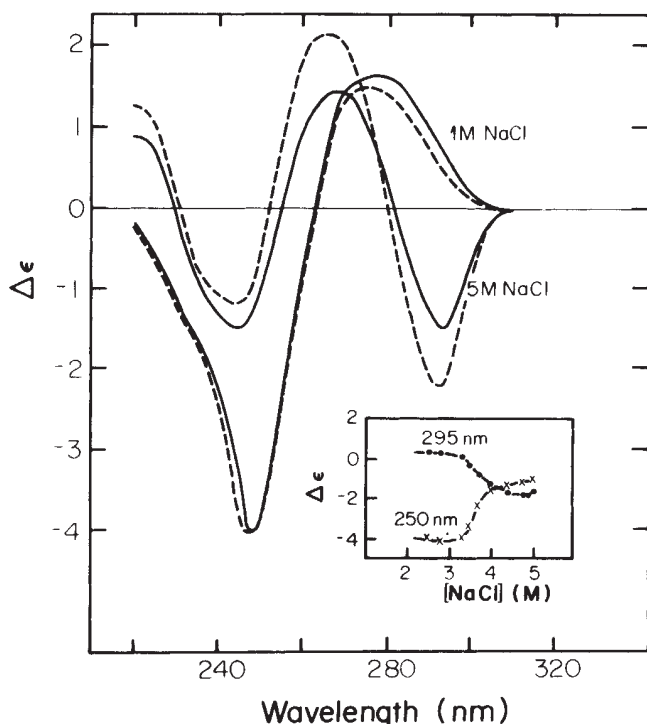


Fig. 2 Circular dichroism spectra of the 157-base pair fragment. The solid lines show the spectra of the 157-base pair fragment in 1.0 and 5.0 M NaCl. The dotted lines show the calculated spectra for a mixture of the *lac* 95-base pair fragment and $(dG-dC)_n \cdot (dG-dC)_n$ at the same molar ratio as in the 157-base pair fragment. The 1.0 and 5.0 M NaCl spectra for the *lac* 95-base pair fragment as well as for the $(dG-dC)_n \cdot (dG-dC)_n$ were taken separately. A solution of $(dG-dC)_n \cdot (dG-dC)_n$ was equilibrated for 4 h at room temperature before the spectra were determined. The extinction coefficients used were the following: *lac* 95-base pair fragment, 6.5×10^3 ; $(dG-dC)_n \cdot (dG-dC)_n$, 7.1×10^3 ; the 157-base pair fragment, 6.7×10^3 . The inset shows the influence of salt concentration on the CD at 295 and 250 nm. The spectra at the different salt concentrations were obtained by diluting the 5.0 M NaCl solution of the 157-base pair fragment with 10 mM cacodylate buffer (pH 7.2) and 0.1 mM EDTA. After each dilution, the sample was equilibrated for 0.5 h at room temperature before the CD spectra were determined. The equilibration time for this fragment was very short; 95% of the transition was obtained after 7 min. $(dG-dC)_n \cdot (dG-dC)_n$ was synthesized as described previously^{2,10}. The *lac* 95-base pair fragment was also prepared and characterized as described previously²⁵ (gift of W. Hillen). All details regarding the CD measurements are described elsewhere²⁵.

methods of cloning^{19,20} and characterizing²¹ other types of DNA polymers with simple sequences, was designed to allow the study of several important questions regarding the Z-structure, including the following: (1) Can Z-DNA exist between flanking segments of B-DNA? (2) What length of $(dG-dC)$ is required to provide a stable Z-conformation? (3) What are the nature and length of the junction between the Z-DNA segments and the flanking B-regions? (4) To what extent does the left-handed Z-type structure influence the biological and physicochemical properties of the neighbouring sequence of random DNA (B-structure) or vice versa? (5) As a Z-type conformation is left handed, does a B $\rightarrow$ Z transition influence the tertiary structure (supercoil) of a plasmid molecule, and will the presence of supercoiling alter the environmental conditions necessary to stabilize a Z-type conformation? (6) Do some biological molecules, such as proteins or small ligands (drugs or carcinogens), specifically interact with the Z-form?

$(dG-dC)_n$ in a plasmid and restriction fragments

Figure 1 illustrates several of the molecules used in this study. A recombinant plasmid (pRW751) was constructed containing an insert 157 base pairs long which was generated by ligating the 95-base pair *lac* operator-promoter fragment²² with a mixture of oligomers prepared by digestion of $(dC-dG)_n \cdot (dC-dG)_n$ with *Fnu*DII. This mixture was ligated into the 'filled-in' *Bam*HI site

of pBR322 by the general procedures described previously^{1,20-22}. The preparation and characterization of this family of recombinant DNAs will be described in detail elsewhere (J. K. and R. D. W., in preparation). A large number of related plasmids were isolated and characterized, many of which were unstable; deletions were observed in the $(dC-dG)$ segments for plasmids which contained $(dG-dC)$ tracts longer than about 40 base pairs (discussed below), but stability was achieved with shorter $(dG-dC)$ tracts. One such plasmid (pRW751) was chosen for further study. It has the structure shown in Fig. 1 and a 157-base pair fragment, excised from it by *Bam*HI and sequenced by methods described previously²³, was shown to consist of the 95-base pair *lac* fragment flanked by strictly alternating dC and dG sequences.

Figure 1 also shows the structure of a polymer of the 157-base pair fragment. This polymer was prepared to internalize the $(dC-dG)$ segments and so evaluate the role of free ends in any possible transition observed in the 157-base pair fragment. The presence of the *Bam*HI sites on the termini of the 157-base pair fragment allowed the formation of this polymer. The polymer was characterized by *Hpa*II digestion followed by quantification of the 'inside-out' 157-base pair fragment relative to the amount of 'half molecules' which are derived from the polymer termini (see below).

Since the first 50-mer of the 157-base pair fragment still contained a small amount of $(dC-dG)$ sequence at the termini, it could not be used to determine rigorously whether the structural transition from a B- to a Z-type structure had to be initiated by $(dC-dG)$ segments at the ends of a DNA molecule or whether it was possible for a segment of $(dC-dG)$ in the interior of a natural sequence to undergo the transition. We therefore produced the inside-out 157-base pair fragment shown in Fig. 1 by *Hpa*II digestion of the 157-base pair 50-mer. This molecule was purified by HPLC on RPC-5 (ref. 24) which cleanly separated the 157-base pair fragment from the small amount of half molecules. Hence, with the inside-out 157-base pair fragment,

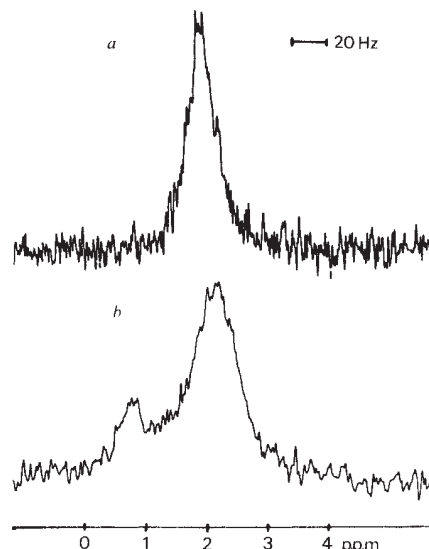


Fig. 3 36.44 MHz ^{31}P -NMR spectra of the 157-base pair fragment in 0.5 and 5.0 M NaCl. The spectra represent a fast Fourier transformation of 4,096 time domain points accumulated following a 90° observing pulse and a 4.2-s delay. Proton resonances were fully and continually decoupled. Exponential multiplication resulting in 0.5 Hz line broadening was applied. Integration was done automatically from manually selected regions of the line integral display. The displayed signals are upfield relative to external 0.95 M H_3PO_4 ; no corrections for salt effects on the deuterium lock signal were made. Temperature was $25 \pm 0.2^\circ\text{C}$. a, 2 mg 157-base pair fragment per ml in D_2O , 0.5 M NaCl, 10 mM cacodylate buffer pH 7.6 and 0.1 mM EDTA; 14,000 accumulations. b, 2 mg 157-base pair fragment per ml in D_2O , 5.0 M NaCl, 10 mM cacodylate buffer pH 7.6 and 0.1 mM EDTA; 19,000 accumulations. The area of the downfield and smaller signal is $\sim 20\%$ of the total resonances. The greater line width at high salt in the 157-base pair fragment could be due to the larger viscosity of the concentrated salt solution, paramagnetic impurities or intermediate phosphorous conformations.

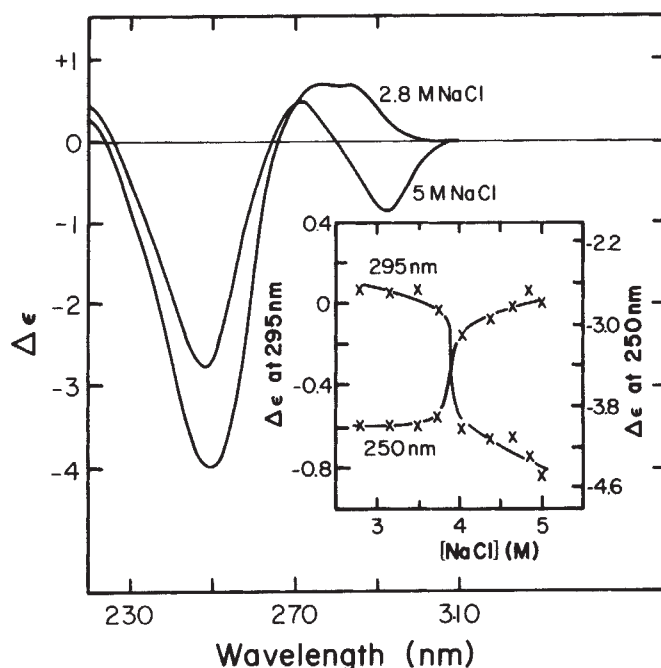


Fig. 4 Circular dichroism spectra of the inside-out 157-base pair fragment in 2.8 and 5.0 M NaCl. The inset shows the CD changes observed at 250 and 295 nm as a function of salt concentration. Other details are described in the legends to Figs 1 and 2.

no (dC-dG) segments are on the termini. It is actually a random mixture of three molecules 138, 157 and 176 base pairs in length which is generated by the *in vitro* ligation used in forming the 50-mer. Note that the (dG-dC) segment in the inside-out 157-base pair fragment contains the *Bam*HI site in its interior.

These three fragments contain identical sequences but in different arrangements. One additional derivative of pRW751 that we have studied is a 220-base pair fragment consisting of the 157-base pair fragment, flanked by segments of pBR322 DNA, 37 and 26 base pairs long (Fig. 1). Hence, the (dC-dG) sequence is flanked by natural sequences which are longer than those for the inside-out 157-base pair fragment. Studies on this 220-base pair fragment supplemented studies with the 50-mer and the inside-out 157-base pair fragment designed to determine if the GATC sequence would adopt a Z-type structure or would remain in a B-conformation in high salt. The 220-base pair fragment differs from the other two DNAs in having (dC-dG) tracts which are uninterrupted by GATC.

Z-type DNA in the 157-base pair fragment

Figure 2 shows the spectra for the 157-base pair fragment in 1.0 and 5.0 M NaCl. In addition, the calculated spectra are shown in identical conditions for the mixture of the appropriate molar ratio of (dC-dG)_n · (dC-dG)_n and the purified 95-base pair fragment. In 1.0 M NaCl, the 157-base pair fragment shows a typical spectrum observed for DNA restriction fragments²⁵; there is very close agreement between the observed spectrum and the spectrum calculated for the component parts. However, in 5.0 M NaCl, the spectrum changes markedly, with negative CD bands at 293 and 245 nm and a positive band at 268 nm. This spectrum agrees well in shape with a calculated spectrum but is lower in amplitude at the peaks at 293 and 268 nm (but not at 245 nm). This general shape in high salt is a result of a marked change in the CD spectrum for (dG-dC)_n · (dG-dC)_n (refs 4-7). The discrepancy between the observed CD curve at 5.0 M NaCl and the calculated curve indicates that a portion of the (dC-dG) arms does not undergo the salt-induced transition to the Z-type conformation (discussed below). Also, we conclude that most (perhaps all) of the 95-base pair *lac* segment remains in a B-structure.

The inset in Fig. 2 shows the cooperative nature of the change in CD at 295 and 250 nm as expected for the salt-induced B → Z

change in conformation in the (dC-dG) arms. In contrast, no cooperative change is observed in similar conditions for the 95-base pair *lac* fragment alone. The midpoint of this transition (~3.6 M) is substantially higher than for (dG-dC)_n · (dG-dC)_n (~2.5 M)⁴, presumably due to the junction with the *lac* sequence.

³¹P-NMR studies were performed to confirm that the inversion in the CD spectrum was due to formation of a Z-type conformation in the (dC-dG) ends of the 157-base pair fragment. Figure 3 shows the spectra observed in 0.5 and 5.0 M NaCl. A single symmetrical peak is observed in the lower salt whereas a second resonance, which is 1.45 p.p.m. downfield from the main resonance, was observed in 5.0 M salt. This result is clearly consistent with previous observations^{16,17}, which documented a splitting of similar magnitude in the phosphorous spectrum of the high salt form of (dG-dC)_n · (dG-dC)_n oligomers.

Integration of the peaks in Fig. 3b indicates that approximately 20% of the phosphorous resonances are in the second small peak which was shifted 1.45 p.p.m. As the (dG-dC) segment comprises ~37% of the 157-base pair fragment, this suggests that virtually all the (dC-dG) tracts have converted to a Z-type conformation. The finding that the NMR studies indicate transition of a larger percentage of the (dC-dG) sequences than is deduced from the CD measurements, may be due to the fact that different conformational features of the molecule are being investigated.

Laser Raman spectroscopy studies (in collaboration with R. M. Wartell) are in complete agreement with these results.

Z-type DNA between flanking B-DNA segments

Figure 4 shows the spectra observed in 2.8 and 5.0 M NaCl for the inside-out 157-base pair fragment. This molecule is of interest because the (dG-dC) segments are totally embodied within the 157-base pair fragment; there are no (dG-dC) segments at the termini.

A typical DNA spectrum²⁵ is observed for this fragment in 2.8 M NaCl. However, in 5.0 M salt the spectrum changes substantially. The trend of the changes is in agreement with the changes found for the 157-base pair fragment (Fig. 2) but their magnitude is less than observed for the 157-base pair fragment and less than calculated for the mixture of the 95-base pair *lac* fragment and the (dC-dG) arms. Thus, the two junctions of the (dG-dC) segments (at the *lac* fragment junctions and at the *Bam*HI site) clearly diminish the capacity of the (dG-dC) segment to undergo the B → Z conformational change. The inset in Fig. 4 shows that the observed transitions are cooperative, as expected for the B → Z conformational transition.

Two other fragments were studied which contained (dC-dG) segments between two naturally occurring B-DNA segments. The polymer containing an average of 50 tandem copies of the 157-base pair fragment (Fig. 1) was also studied by CD. This fragment differs from the inside-out 157-base pair molecule in length but not sequence. The CD spectra of this polymer at 2.8 and 5.0 M NaCl (data not shown) are almost identical to those shown in Fig. 4 for the inside-out 157-base pair fragment. Also the transition is cooperative, with a midpoint which is identical to that found for the inside-out 157-base pair fragment (3.9-4.0 M). Thus, the length of the molecule has no apparent influence on these properties.

Moreover, the CD properties of the 220-base pair fragment (Fig. 1) show the same trends as observed for the inside-out 157-base pair fragment and for the polymer of the 157-base pair fragment, but the magnitude of the change is somewhat less, as expected from its smaller (dC-dG) content. The quantitation of these transitions and their interpretation regarding DNA structure are described below. Because the same type of CD transition is observed with the 220-base pair fragment, we conclude that the transition is not a function of some peculiarity of the other three fragments (157-base pair, polymer of 157-base pair,

and inside-out 157-base pair fragments) but is a general property of the (dC-dG) sequence embodied in a natural DNA molecule.

Junction between B- and Z-type conformations

It is possible to calculate the fraction of the (dC-dG) segments which are in a Z-type conformation for all the fragments studied (Fig. 1), as virtually no CD signal at 295 nm is observed in 5.0 M NaCl for the *lac* 95-base pair fragment. Hence, any negative signal observed for the fragments at 295 nm must be due to (dC-dG) segments which are transformed into the Z-type conformation. Thus, by comparison with the calculated CD signal for the component parts, we can determine the amount of Z-type structure. For these calculations, we shall use an average of 29 base pairs for the (dC-dG) tracts (average of 26 and 32). For the 157-base pair fragments, we deduce that 65% (or 19 out of an average of 29 base pairs in each segment) are in a Z-type conformation. As each (dC-dG) segment is joined only once to the *lac* 95-base pair segment, 10 base pairs are in a conformation other than Z-type. We believe that these are at the junction of the (dC-dG) segment and the *lac* 95-base pair fragment.

For the inside-out 157-base pair fragment, 36% (or 10 base pairs out of an average length of 29) are in a Z-type conformation. As the (dG-dC) segment interfaces at both ends with a random type sequence, $19/2 = 9.5$ base pairs (or 10 base pairs) constitute the junction region. This, of course, presumes that the GATC *Bam*HI site disrupts the Z-type conformation of the (dG-dC) tract.

For the polymer of the 157-base pair fragment, 14% (or 4 base pairs out of an average of 29) are in a Z-type conformation. As the (dC-dG) region in this polymer also interfaces at both ends with random DNA sequences, $25/2 = 12.5$ base pairs constitute each of the junction regions which are not in a Z-type conformation. Similarly, for the 220-base pair fragment, 16% (5 base pairs out of an average of 29) are in a Z-type conformation, leaving a junction region of 12 base pairs.

Thus, the length of the junction regions for all four molecules ranges from 9.5 to 12.5 base pairs. The remarkable consistency in the length of this region implies the necessity for a transition region of ~ 11 base pairs from a B-conformation to the Z-type structure. The structure of this intermediate region (approximately one turn of helix) is unknown. Wing *et al.*²⁶ have determined the crystal structure of d(CGCGAATTCGCG); their finding that the CGCG termini are not in a Z-structure is consistent with our results.

These calculations make several assumptions, including the following: (1) The CD signal from the segments of *lac* and pBR322 DNA in these restriction fragments is the same as for the 95-base pair segment alone. That is, the conformation of the natural sequences as judged by CD is not influenced by their junction to Z-type DNA. (2) The GATC which comprises the *Bam*HI site is not in a Z-type conformation when the surrounding (dC-dG) tracts are in a Z-type structure for the cases of the polymer of the 157 base pairs and the inside-out 157-base pair fragment. (3) The junction regions (9.5–12.5 base pairs in length) of (dC-dG) residues which are not in a Z-type conformation give no CD signal at 295 nm.

Z-DNA in a supercoiled plasmid

It was of interest to determine if a Z-type conformation can exist within a covalently closed and circular DNA. On the basis of previous studies on supercoiled DNA, it is logical that a B $\rightarrow$ Z transition in a (dC-dG) segment should alter the number of supercoil turns in a plasmid DNA because the supercoil number is dictated by the specific linking difference which is sensitive to conformational changes in the primary helix (reviewed in refs 27, 28).

pRW751 was a suitable molecule for studying this question. The total length of (dC-dG) sequences in pRW751 is 58 base

pairs or 1.3% of the length of the total recombinant plasmid (4,516 base pairs). The unwinding of one turn of primary helix removes one right-hand supercoil. In addition, each turn of left-handed primary helix which is induced will also remove one right-hand supercoil. Thus, we can calculate that a complete transition of 58 base pairs of (dC-dG) from B $\rightarrow$ Z would theoretically remove 10.3 supercoils from pRW751 ($58/10.4$ base pairs per turn + $58/12$ base pairs per turn = 10.3).

We have established conditions for running agarose gels in NaCl solutions up to 5.0 M (Fig. 5) to determine the number of supercoil turns in a DNA. pRW451 served as a control for these electrophoresis experiments as it is identical to pRW751 apart from containing a 174-base pair insert (formerly a *Hha*I fragment of pBR322 DNA which contains 54% G + C) instead of the 157-base pair insert in pRW751. This 174-base pair insert does not contain unusual sequence features or stretches of (dC-dG) sequence²⁹.

Figure 5 shows the gel electrophoresis patterns observed for pRW451 and pRW751 in 3.0 and 4.5 M NaCl. In 3.0 M, the partially relaxed family of topoisomers for both DNAs comigrate (lanes a and b). For both DNAs in these electrophoresis conditions, the family ranges from approximately five right-

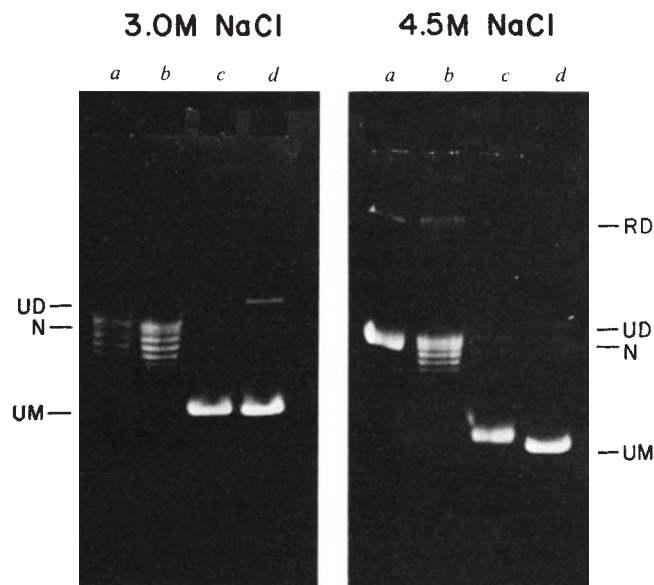


Fig. 5 High salt gels on supercoiled plasmids. Monomeric supercoiled pRW751 was separated from higher forms (such as dimer, tetramer) by HPLC on RPC-5 using a 1.5 mm $\times$ 20-cm column at 43 °C. Elution was with a 0.65–0.8 M KCl gradient using the general techniques described previously²⁴. pRW451 contains the 174-base pair *Hha*I fragment from pBR322 cloned into the filled-in *Bam*HI site of pBR322 (G. D. Stafford and R. D. W., unpublished). pRW451 was isolated in a similar manner. Partially relaxed pRW751 and pRW451 were prepared by treatment with wheat-germ topoisomerase (gift of R. R. Burgess). Reaction mixtures (100 μ l) contained 20 μ g DNA and 8 μ g enzyme in 50 mM Tris-HCl pH 7.6, 0.1 mM dithiothreitol, 10 mM MgCl₂ and 2.5 mM spermidine. The reaction was for 30 min at 37 °C which gives a terminal reaction product of right-handed supercoiled topoisomers. The unrelaxed and partially relaxed plasmids were electrophoresed through 1% agarose gels with a buffer of 80 mM Tris-HCl, 40 mM sodium acetate, 4.4 mM disodium EDTA and 0.525 M acetic acid, all at pH 8.3, plus either 3.0 or 4.5 M NaCl. The gels were 8 cm $\times$ 14 cm $\times$ 1 mm. The DNA samples were equilibrated in a loading buffer identical to the electrophoresis buffer but with an identical molarity of NaClO₄ substituted for NaCl and 0.05% bromophenol blue added. The gels were electrophoresed for 36 h at 25 V for the 4.5 M NaCl gel and 20 V for the 3 M NaCl gel. The gels were desalted by soaking in water for 10 min and then stained with 100 μ g ml⁻¹ ethidium bromide. Abbreviations are: UM, unrelaxed monomer; N, nicked monomer; UD, unrelaxed dimer; RD, relaxed dimer. The following samples were electrophoresed in the designated lanes: a, partially relaxed pRW751; b, partially relaxed pRW451; c, unrelaxed pRW751; d, unrelaxed pRW451. Resolution of fully relaxed plasmid from supercoiled DNA containing +1 or -1 supercoils is difficult in these conditions. Similarly, the logarithmic nature of topoisomer migration prevents the resolution of individual bands for topoisomers containing ~ 10 supercoils.

handed supercoil turns to fully relaxed DNA (the exact number of supercoil twists is uncertain due to our inability to resolve topoisomers near the fully relaxed position, but we believe that this number is accurate within at least ± 1). This experiment was performed with plasmids which were partially relaxed with topoisomerase to obtain better resolution of the family of topoisomers and thus enhance our ability to detect smaller changes in supercoil number. Figure 5, lanes *c* and *d*, shows that the unrelaxed plasmids co-migrate in 3.0 M NaCl.

When similar studies were done in 4.5 M NaCl, no change was observed for the migration pattern of the partially relaxed control plasmid (pRW451) (Fig. 5, lane *b*). However, Fig. 5, lane *a*, shows that all the topoisomers for pRW751 co-migrated in these conditions at the fully relaxed position. Thus, a loss of as many as five supercoil turns was observed, but there was no apparent generation of left-handed supercoil twists.

This striking result clearly demonstrates that a salt-induced structural change occurred in pRW751 which had a dramatic influence on the supercoil number, and is consistent with a conformational transition from B- to a Z-type structure for at least a portion of the (dC-dG) tracts in pRW751.

A small difference was observed in the migration of pRW451 and pRW751 at 4.5 M NaCl (Fig. 5, lanes *c*, *d*). Based on our experience with topoisomer migration patterns, we believe this difference is equivalent to a loss of four to eight supercoil turns. However, the relatively poor resolution of topoisomers of high superhelical density makes precise quantification difficult, especially in these electrophoresis conditions.

Electrophoresis studies were performed on the partially relaxed plasmids at several salt concentrations between 3.5 and 4.0 M in addition to those shown in Fig. 5 to determine the degree of cooperativity. At salt concentrations of 3.5 M to at least 4.0 M, the topoisomers with more supercoil turns decreased in supercoil number at lower salt concentrations than topoisomers with fewer supercoil turns, suggesting that supercoiling contributes a favourable free energy to the B- to Z-type transition.

Thus, the studies with both the partially relaxed and unrelaxed plasmids indicate an unwinding of about five supercoil turns (conceivably not fewer than four and not more than eight) in 4.5 M NaCl. In the calculations described above, if the entire (dC-dG) segment in pRW751 converted from a B $\rightarrow$ Z conformation, we should have observed a loss of 10.3 supercoil turns. As this large an effect was not observed, the entire length of the (dC-dG) segments may not be in a Z-type structure, as predicted from the CD results.

Conclusions

Circular dichroism and ^{31}P -NMR studies on fragments show that the (dC-dG) segments undergo a cooperative transition to a left-handed conformation in the presence of high salt. The natural sequences from the *lac* operon and pBR322 do not appear to convert to a left-handed conformation. In all cases a junction of ~ 11 base pairs seems to exist between the left-handed conformation and B-DNA which is in neither a Z-type nor a B-conformation but must be intermediate in nature.

Moreover, electrophoretic analyses of plasmids containing the (dG-dC) sequence demonstrate a cooperative transition from a right-handed conformation to a left-handed Z-type structure. The presence of the (dG-dC) segment in the amount of 1.3% of the recombinant plasmid is sufficient to have a profound influence on the supercoil properties of the DNA. These observations are consistent with prior determinations on the extreme flexibility of the DNA chain (see ref. 30, reviewed in ref. 1).

Recombinant plasmids related to pRW751 commonly showed deletions in the (dG-dC) tracts, a type of behaviour unprecedented in our experience with a substantial variety of cloning studies^{1,19-22,24,31-33}. It seems that (dG-dC) segments longer than ~ 40 base pairs may not be stable in *Escherichia coli* (J. K. and R. D. W., unpublished).

Biological implications of left-handed DNA

The demonstration that different segments of one DNA chain can co-exist with conformations as different as a right-handed B-helix and a left-handed Z-type structure may have profound implications for the role of DNA structure in regulatory processes. Whereas this is not the first implication of DNA conformations and properties as important components in gene regulation or of left-handed DNA structures (reviewed in ref. 1), it is the first clear demonstration of left-handed DNA which is adopted by a naturally replicating molecule and by portions of restriction fragments isolated from that plasmid.

A segment of left-handed DNA in a viral or cellular DNA which is principally comprised of right-handed conformations would be an easily recognizable locus for binding of specific proteins or ligands³⁴ due to the dramatic difference in conformation. Moreover, the short region (approximately one turn of helix) at the interface between the B- and Z-conformations probably consists of intermediate structures which would also provide unique interaction contacts for ligand binding.

Our studies demonstrate that changes in the conformation of the primary helix in as little as 1.3% of pRW751 can profoundly influence its supercoil structure. Thus, conformational transitions in tiny portions of a chromosome can have large effects on the overall topography of a genome. As the extent of supercoiling has differential effects on the transcription properties of different genes (reviewed in ref. 1), this implicates conformational transitions in the primary helix as a genetic regulatory mechanism.

Our results clearly demonstrate that the cooperative salt-induced transition can be initiated within a polynucleotide chain. It is possible that a smooth and nondisruptive transition may be sterically and energetically feasible between B- and Z-conformations, as shown³⁵⁻³⁷ for the interface of B- and A-structures.

It is uncertain if appropriate natural sequences, in fact, undergo a transition to a Z-structure *in vivo*. However, several microorganisms (for example, halobacteria) as well as crustacea have high intracellular salt concentrations. Moreover, agents other than high salt may effect the B $\rightarrow$ Z transition *in vivo*.

Do sequences exist in viral or cellular DNAs which can adopt a left-handed conformation? We are unaware that a long tract of perfect (dG-dC) has yet been found in a duplex DNA. However, 7-base pair lengths of this sequence are found in a hairpin region of four related parvovirus genomes (reviewed in ref. 1). Moreover, a number of related repeating dinucleotide sequences have been found³⁸⁻⁴³.

This work was supported by NIH grant CA 20279 and NSF grants PCM 77-15033 and PCM 77-19927.

Received 3 November 1980; accepted 2 February 1981.

- Wells, R. D. *et al. Prog. Nucleic Acid Res. molec. Biol.* **24**, 167-267 (1980).
- Wells, R. D., Larson, J. E., Grant, R. C., Shortle, B. E. & Cantor, C. R. *J. molec. Biol.* **54**, 465-497 (1970).
- Wells, R. D. *et al. Crit. Rev. Biochem.* **4**, 305-340 (1977).
- Pohl, F. M. & Jovin, T. M. *J. molec. Biol.* **67**, 375-396 (1972).
- Pohl, F. M., Jovin, T. M., Baehr, W. & Holbrook, J. J. *Proc. natn. Acad. Sci. U.S.A.* **69**, 3805-3809 (1972).
- Pohl, F. M., Ranade, A. & Stockburger, M. *Biochim. biophys. Acta* **335**, 85-92 (1973).
- Pohl, F. M. *Nature* **260**, 365-366 (1976).
- Grant, R. C., Harwood, S. J. & Wells, R. D. *J. Am. chem. Soc.* **90**, 4474-4476 (1968).
- Mitsui, Y. *et al. Nature* **228**, 1166-1169 (1970).
- Grant, R. C., Kodama, M. & Wells, R. D. *Biochemistry* **11**, 805-815 (1972).
- Wang, A. H.-J. *et al. Nature* **282**, 680-686 (1979).
- Crawford, J. L. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 4016-4020 (1980).
- Wang, A. H.-J. *et al. Science* **211**, 171-176 (1980).
- Crawford, J. L. *et al. Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Drew, H., Takano, T., Tanaka, S., Itakura, K. & Dickerson, R. E. *Nature* **286**, 567-573 (1980).
- Patel, D. J., Canuel, L. S. & Pohl, F. M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 2508-2511 (1979).
- Simpson, R. T. & Shindo, H. *Nucleic Acids Res.* **8**, 2093-2103 (1980).
- Arnott, S., Chandrasekaran, R., Birdsall, D. L., Leslie, A. G. W. & Ratliff, R. L. *Nature* **283**, 743-745 (1980).
- Klein, R. D., Selsing, E. & Wells, R. D. *XIII int. Congr. Biochem.* 01-1-H26 (1979).
- Selsing, E. & Wells, R. D. *Nucleic Acids Res.* **6**, 3025-3040 (1979).
- Klein, R. D., Selsing, E. & Wells, R. D. *Plasmid* **3**, 88-91 (1980).
- Hardies, S. C. *et al. J. Biol. Chem.* **254**, 5527-5534 (1979).
- Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560-564 (1977).
- Wells, R. D. *et al. Meth. Enzym.* **65**, 327-347 (1980).
- Hillen, W. & Wells, R. D. *Nucleic Acids Res.* **8**, 5427-5444 (1980).

26. Wing, R. *et al.* *Nature* **287**, 755–758 (1980).
27. Bauer, W. R. A. *Rev. Biophys. Bioengng* **7**, 287–313 (1978).
28. Wang, J. C. *Trends biochem. Sci.* **5**, 219–221 (1980).
29. Sutcliffe, J. G. *Cold Spring Harb. Symp. quant. Biol.* **43**, 77–90 (1978).
30. Wartell, R. M., Larson, J. E. & Wells, R. D. *J. biol. Chem.* **250**, 2698–2702 (1975).
31. Hardies, S. C. & Wells, R. D. *Gene* **7**, 1–14 (1979).
32. Panayotatos, N. & Wells, R. D. *J. molec. Biol.* **135**, 91–109 (1979); *Nature* **280**, 35–39 (1979).
33. Müller, U. R. & Wells, R. D. *J. molec. Biol.* **141**, 1–24 (1980); *J. molec. Biol.* **141**, 25–41 (1980).
34. Sage, E. & Leng, M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4597–4601 (1980).
35. Selsing, E., Wells, R. D., Early, T. A. & Kearns, D. R. *Nature* **275**, 249–250 (1978).
36. Selsing, E. & Wells, R. D. *J. biol. Chem.* **254**, 5410–5416 (1979).
37. Selsing, E., Wells, R. D., Alden, C. J. & Arnott, S. *J. biol. Chem.* **245**, 5417–5422 (1979).
38. Nishioka, Y. & Leder, P. *J. biol. Chem.* **255**, 3691–3694 (1980).
39. Sures, I., Lowry, J. & Kedes, L. H. *Cell* **15**, 1033–1044 (1978).
40. Maio, J. J. *Handbk Biochem.* **2**, 379–404 (1976).
41. Kimmel, A. R. & Firtel, R. A. *Cell* **16**, 787–796 (1979).
42. Schachman, H. K., Adler, J., Radding, C. M., Lehman, I. R. & Kornberg, A. *J. biol. Chem.* **235**, 3242–3249 (1960).
43. Burd, J. F. & Wells, R. D. *J. molec. Biol.* **53**, 435–459 (1970).

Evolution of the 87A and 87C heat-shock loci in *Drosophila*

A. J. Leigh Brown & D. Ish-Horowicz

Imperial Cancer Research Fund, Mill Hill Laboratories, Burtonhole Lane, London NW7 1AD, UK

Several species related to Drosophila melanogaster have two loci containing sequences which code for the major heat-shock protein of molecular weight 70,000 (hsp70). These sequences are arranged as two pairs of inverted repeats. The structure of the 87C locus in D. melanogaster seems to have arisen through DNA insertion. Hsp70-coding sequences at the two loci are not evolving independently. Gene conversion between hsp70 genes has apparently occurred both within and between loci.

In higher eukaryotes, a single enzymatic function is frequently encoded at more than one locus. The duplicated loci need not be closely linked but are likely to derive from a single ancestral locus. A second form of gene reiteration is the presence of tandem arrays of repeated genes, for example ribosomal DNA and 5S genes. We have studied the evolution in *Drosophila* of a gene system showing both styles of duplication for which detailed molecular analysis is possible. Several features have emerged which suggest that classical views on mutation and evolution in such systems are inadequate.

Exposure of *Drosophila* tissue to a temperature of 37 °C activates a characteristic set of about nine chromosomal puffs and also induces the synthesis of a new set of about nine proteins (reviewed in ref. 1). As the heat shock inhibits previous RNA and protein synthesis, the resulting pattern of macromolecular synthesis is greatly simplified. Several laboratories have used heat-shock mRNA to select cloned genomic DNA fragments that encode the major 70,000-molecular weight heat-shock protein (hsp70). Such DNA fragments have been localized to two heat-induced puffs (87A and 87C)^{2–4}. This has been confirmed by genetic analysis using mutations that delete one or other locus. In this way, the two hsp70-coding loci were mapped to the sibling bands 87A7 and 87C1 (ref. 5). The sibling species

Drosophila simulans also makes a 70,000-molecular weight heat-induced protein⁶, and *in situ* hybridization shows that it too is encoded at two loci, 87A and 87C (ref. 2). A second type of heat-induced sequence, called 'αβ' sequences, is found at 87C1 but not at 87A7 in *D. melanogaster*⁷. These sequences seem to be absent at both hsp70 loci in *D. simulans*, although, like *D. melanogaster*, they are found at the chromocentre².

The arrangement of hsp70-coding sequences in *D. melanogaster* has been elucidated by use of cloned DNA sequences. The total number of hsp70 genes varies from five to about eight according to genotype, there being two copies at 87A7 and three to five at 87C1 (refs 8–10). (In this discussion, 'gene' refers to an mRNA-coding element and adjacent conserved sequences. 'Locus' designates the cytogenetic origin, that is, 87A7 or 87C1.) Each of the hsp70 genes seems to be organized within a 2.5-kilobase conserved element (*z*), comprising the 2.1-kilobase mRNA-coding region (*z*_c) and the 400 5'-terminal bases that are not transcribed (*z*_{nc})^{3,8,9,11}. Nevertheless, the *z* elements at 87A7 and 87C1 differ by some characteristic restriction sites^{8,10}.

Studies on gene arrangement at the hsp70 loci have shown that the two hsp70 genes at 87A7 are about 1.7 kilobases apart (Fig. 1) in an opposite orientation^{9,12,13}. The copies at 87C1 are

Table 1 Comparison of 87A and 87C heat-shock loci in *D. melanogaster*, *D. simulans* and *D. mauritiana*

	Total no. of restriction sites used in comparison	Sites	Restriction sites in common and nucleotide substitution per site (δ)			
			With <i>D. melanogaster</i>		With <i>D. simulans</i>	
			δ	Var (δ)	δ	Var (δ)
<i>D. melanogaster</i>						
Protein coding	26					
Non-coding	39 (<i>D. mauritiana</i>)					
	37 (<i>D. simulans</i>)					
<i>D. simulans</i>						
Protein coding	24	22	0.0213	0.000152		
Non-coding	38 (<i>D. mauritiana</i>)	31	0.0295	0.000145		
	37 (<i>D. melanogaster</i>)					
<i>D. mauritiana</i>						
Protein coding	24	22	0.213	0.000152	24	0.00
Non-coding	34 (<i>D. melanogaster</i>)	30	0.327	0.000165	30	0.0257
	32 (<i>D. simulans</i>)					0.000132

Division of the hsp70 genes into protein-coding and non-protein-coding regions was made from ref. 35. The protein-coding region includes the 3'-terminal *SalI* sites and the 5' *BglII* sites. It does not include the 5' *PstI*, *SalI* or *EcoRI* (*D. mauritiana*) sites. The total number of restriction sites is the number of sites in one species which could have been detected in a second. This number varies in different comparisons because only the restriction sites closest to the *z* elements could be mapped. The mean frequency of nucleotide substitution and its variance (Var δ) were estimated after Nei and Li⁴².

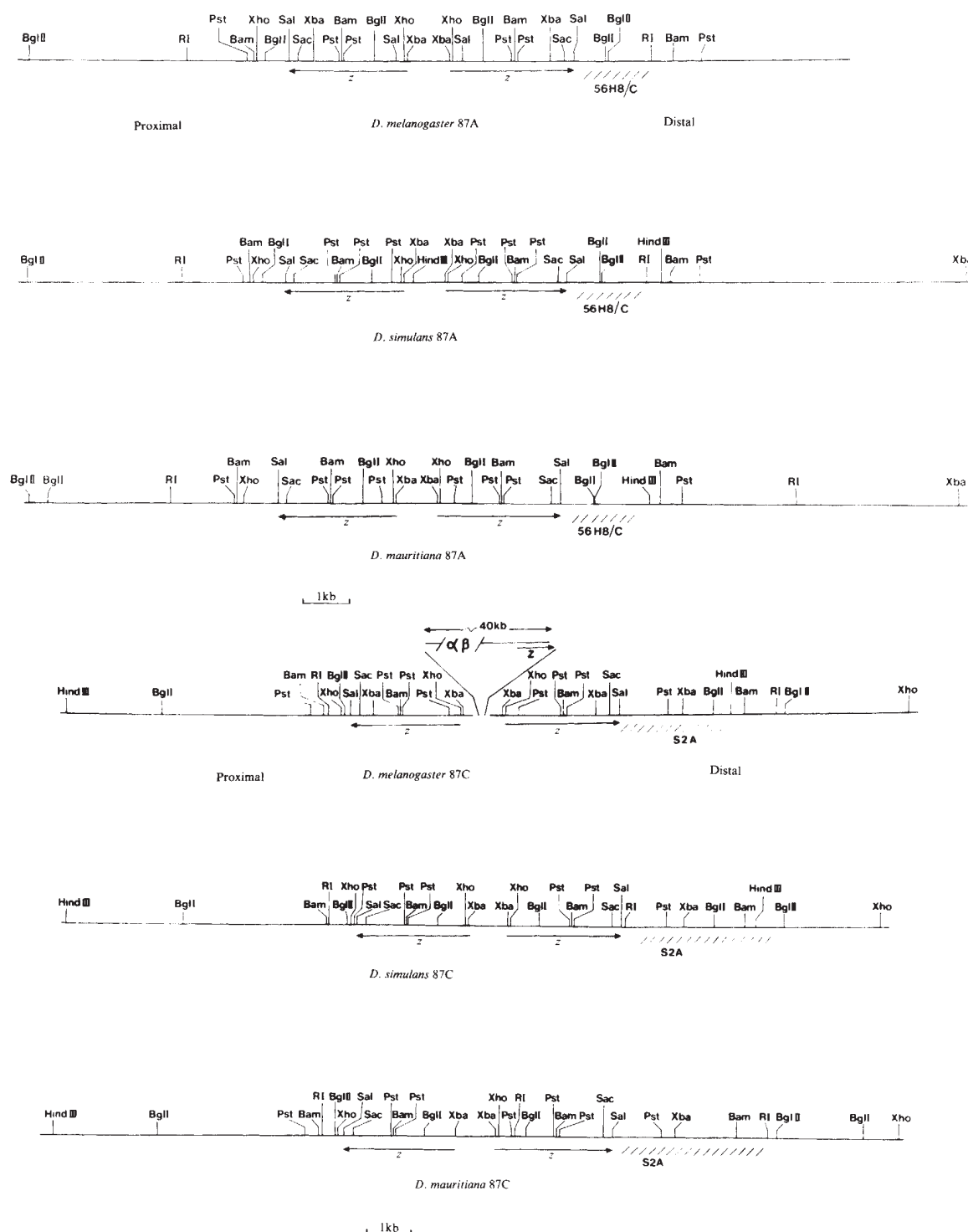


Fig. 1 Organization of hsp70-coding loci. DNA extraction, purification, digestion and electrophoresis were as described previously^{8,13}. DNA was prepared from adult flies of inbred strains of *D. simulans* and *D. mauritiana* obtained by single pair matings for several generations. The DNA was analysed by the technique of Southern¹⁵ as adapted by Wahl *et al.*⁴³. The organization of hsp70-coding loci in *D. melanogaster* is redrawn from ref. 13. (The region depicted in the inset at the 87C1 locus is not drawn to scale.) The hsp70-coding sequences in our inbred strains of *D. simulans* and *D. mauritiana* were positioned and oriented at the cytogenetic loci using two locus-specific probes. Plasmid 56H8/C is derived from plasmid 56H8 (ref. 16) and contains a subcloned *SalI*-*EcoRI* fragment that lies just downstream of the distal hsp70 gene at 87A7. This probe hybridizes only to restriction fragments of *D. melanogaster* known to come from 87A7 (data not shown). It allows us to visualize restriction fragments from 87A7 and to orientate the 87A7 restriction maps on the chromosome. Similarly, plasmid S2A is derived from a cosmid clone cHS2 bearing two tandem genes from the distal region at 87C1 in *D. melanogaster* (A. Udvardy, personal communication). We have subcloned the *SalI*-*EcoRI* fragment immediately downstream from the most distal *z* element of cHS2, making a probe specific for 87C1. The *z* elements whose 3'-adjacent DNA sequences hybridize to these probes are aligned in the figure. Note that the designation of the homologous genes as 'proximal' and 'distal' is inappropriate in *D. simulans* and *D. mauritiana* because of the large inversion on this arm of the third chromosome which is found in these species¹⁴. To avoid confusion only the *D. melanogaster* restriction maps have been thus labelled. kb, Kilobase.

present in two domains, a single proximal sequence separated from tandem distal sequences by about 38 kilobases containing the $\alpha\beta$ heat-induced sequences¹³ (Fig. 1).

In this article we compare the gene organization of hsp70 sequences in three closely related species, *D. melanogaster*, *Drosophila mauritiana* and *D. simulans*, which show very similar polytene chromosome banding patterns¹⁴. We deduce the probable ancestral organization, and use the restriction maps to estimate the frequency of nucleotide substitution between the species. We also find that the hsp70 genes within a species are much more alike than would be expected if they were evolving independently. This suggests that there is gene conversion both within and between hsp70 loci and we discuss a possible mechanism by which this could occur.

Organization of hsp70-coding sequences in *D. simulans* and *D. mauritiana*

Hsp70 genes of *D. simulans* and *D. mauritiana* are found at two loci, 87A and 87C, as in *D. melanogaster* (ref. 2; S. M. Pinchin, personal communication). We have studied their organization in adult fly DNA using the transfer hybridization method of Southern¹⁵. Figure 1 shows the restriction maps of the hsp70 loci of *D. simulans* and *D. mauritiana*. The maps reveal that the hsp70-coding regions lie within a conserved z element, as in *D. melanogaster*. In particular, the *Xba*I and *Xho*I restriction sites characteristic of z_{nc} are conserved, strengthening the idea that z_{nc} is important for hsp70 expression³. In Fig. 2 we show certain restriction digests that establish the conservation of z elements and others that demonstrate the presence of four coding sequences in these species. Our *D. simulans* and *D. mauritiana* stocks did not show the variation in hsp70 gene number that is found in *D. melanogaster*, but we have not examined other independent lines.

Both hsp70 loci of *D. simulans* and *D. mauritiana* have the same organization—two hsp70 genes in opposite orientations (Fig. 1). We have demonstrated the presence of inverted-repeat hsp70 genes by the presence of S_1 nuclease-resistant genes in denatured genomic DNA (Fig. 2C). We find two S_1 -resistant ('snapback') bands of 2.6 and 2.8 kilobases which include the whole z element. [The absence of an *Xho*I site in an 87C gene of *D. mauritiana* is due to a restriction site polymorphism (A.J.L.B., unpublished data).] This orientation is confirmed by digestion of chromosomal DNA with the restriction enzyme *Sac*I which cuts every hsp70 gene near the 3' end^{4,13}. Only a single 5.3-kilobase band is found (Fig. 2C), showing that the genes are paired and the same distance apart.

Evolution of the organization of hsp70 genes in *Drosophila*

The members of the *D. melanogaster* subgroup are closely related. A previous study on the polytene chromosome banding patterns has suggested that *D. simulans* and *D. mauritiana* are more closely related to each other than to *D. melanogaster*¹⁴. Two more distantly related species, *Drosophila teisseri* and *Drosophila yakuba*, also have two pairs of hsp70 genes both arranged as inverted repeats (Figure 2C). This indicates that this was the ancestral organization which existed at both 87A7 and 87C1 before the separation of *D. melanogaster* from the other species.

The 87C1 locus in *D. melanogaster* is very different from all the other hsp70 loci (Fig. 1)¹³. It must have arisen through DNA insertion between the two diverging hsp70 genes of the ancestral locus. The intervening region between hsp70 domains at this locus is very large and contains many repeated sequences, including the α and β elements which are also found at the chromocentre and some other sites⁷. The α and β elements are also present at the chromocentre in *D. simulans*² and so may be presumed to have been present before the speciation of *D. simulans* and *D. melanogaster*. As the chromocentral α and β elements of *D. melanogaster* are non-adjacent¹⁶, it would require at least two events to introduce contiguous $\alpha\beta$ elements at 87C1.

Insertions between the two hsp70 genes at 87A7 seem to be common in *D. melanogaster*. The 56H8 clone of Schedl *et al.*¹⁷ derives from this locus and contains a dispersed repetitive sequence between the two hsp70 genes¹⁸. Other changes in the structure of the 87A7 locus in chromosomes derived from natural populations seem to be due to independent events of this type (A.J.L.B., unpublished observations). After heat shock the inserted $\alpha\beta$ elements are transcribed from 87C1 but no other sites to give an abundant polyadenylated RNA^{7,16}. Thus the sequence rearrangements at 87C1 have resulted in activation of previously unexpressed $\alpha\beta$ sequences, although the transcripts do not seem to be essential for viability or fertility^{13,19}. This suggests that DNA rearrangements during evolution may be responsible for more rapid change in gene function than base substitution.

Nucleotide substitution at the hsp70 loci

We can study the evolution of the hsp70 loci using the restriction maps (Fig. 1) to estimate the rate of nucleotide substitution. This is possible because the three maps are independently derived and the two loci were distinguished using locus-specific probes,

Fig. 2 A, Hsp70 genes are arranged in a z element in *D. mauritiana* and *D. simulans*. Autoradiograms of transferred agarose gels were hybridized with the plasmids pPW232.1 and pPW229.1, probes for the 5' and 3' halves of the hsp70 gene respectively². In lanes *m* are molecular-weight markers of size indicated in kilobases. In lanes *a*–*c*, *D. mauritiana* DNA has been digested with *a*, *Xba*I/*Bam*HI (pPW232.1); *b*, *Xba*I/*Sac*I (pPW232.1); *c*, *Bam*HI/*Sac*I (pPW229.1). B, There are four copies of hsp70-coding sequences in *D. mauritiana* and *D. simulans*. *d*, An *Xba*I/*Bgl*II digest of *D. mauritiana* DNA probed with pPW229.1. *D. simulans* gives identical patterns with all the above digests. C, 'Snapback' hsp70 genes in species related to *D. melanogaster*. Lane *e* contains a *Sac*I digest of *D. mauritiana* DNA (an identical pattern is obtained with *D. simulans*). In lanes *f*–*i*, DNA was denatured, reannealed and digested with S_1 nuclease exactly as described previously¹³, before electrophoresis and transfer hybridization to plasmid pPW229.1. Lane *f* is *D. teisseri*, *g* is *D. yakuba*, *h* is *D. simulans*, and *i* is *D. mauritiana*.

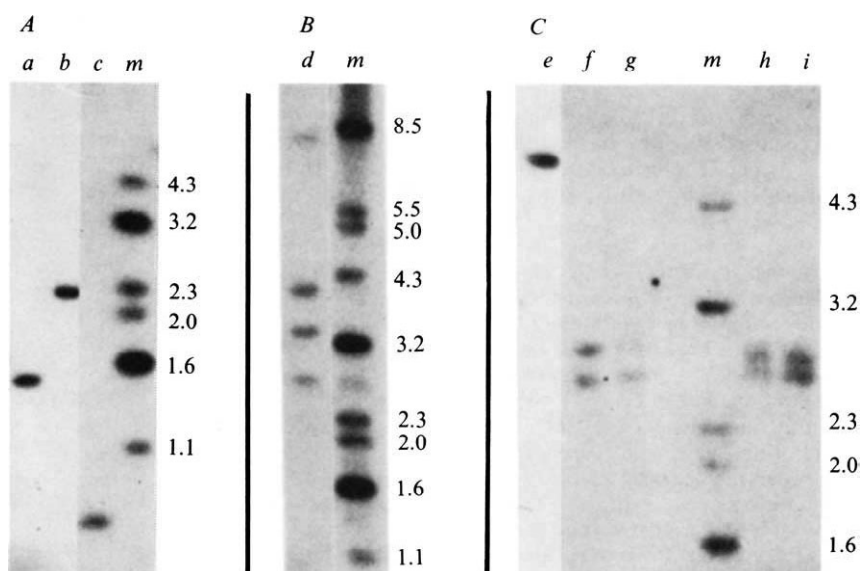
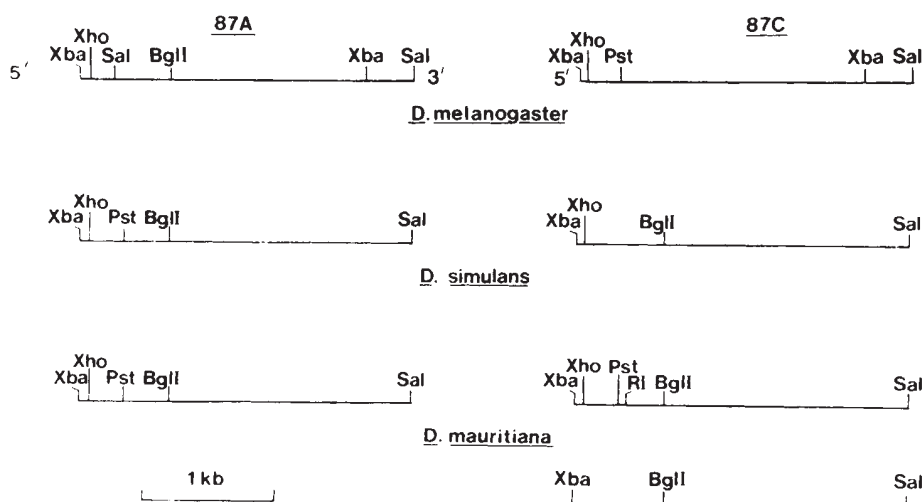


Fig. 3 Restriction site differences between z elements from the 87A and 87C loci in three *Drosophila* species. The 5'-terminal *Xba*I and 3'-terminal *Sal*I sites are indicated and the distribution of the variable sites shown. Within a locus z elements are identical unless otherwise indicated. To estimate the probability of parallel change in a restriction site in both members of a pair of z elements (see text), we calculated the proportion of restriction sites conserved between homologous z elements in *D. melanogaster* and each of the other two species⁴². (Independent evidence suggests that *D. simulans* and *D. mauritiana* are more closely related to each other than either is to *D. melanogaster*¹⁴.) The four proportions were pooled to give an overall estimate of the mean number of restriction sites conserved of 0.828. The probability that a given site is different in two homologous z elements is thus $1 - 0.828 = 0.172$, and the probability that the same change occurred independently in a pair of sequences is $0.0297 (=0.172^2)$. We have found six examples of parallel change in 36 comparisons of restriction site pairs ($P = 0.00062$). kb, Kilobase.



not restriction-site homology. Comparison of hsp70 genes from 87A and 87C in *D. melanogaster* has shown that the non-protein-coding regions have diverged much more than the protein-coding regions^{11,20}. Presumably, the latter are constrained by the need to retain protein function. We have therefore considered the nucleotide substitution of the region coding for hsp70 protein separately from that of the surrounding sequences (Table 1).

We can use the degree of nucleotide divergence to estimate the rate of substitution for the non-protein-coding regions. Carson²¹ has described a correlation in Hawaiian *Drosophila* between estimated divergence time and genetic distance based on frequencies of enzyme variants. The genetic distance between *D. simulans* and *D. melanogaster*²² would suggest a divergence time of about 1×10^6 generations. We estimate the degree of nucleotide substitution in non-coding DNA to be 0.03 per nucleotide (Table 1). This is equivalent to a rate of nucleotide divergence of the order of 3×10^{-8} per nucleotide per generation. Although estimations of divergence time are subject to large errors, we note that this rate is consistent with estimates for primates (approximately 1×10^{-8} per nucleotide per generation) obtained from the measurement of melting points of interspecific hybrid DNA molecules^{23,24}.

Intra-locus gene conversion between z elements

We have remarked previously that z elements from the same locus have similar restriction maps in *D. melanogaster* and suggested that this may imply an intra-locus correction mechanism⁸. We now have supporting data on z elements at six hsp70 loci. There are six examples of parallel change in a particular site in both z elements at the same locus [Fig. 3; *Pst*I (87A *D. melanogaster*); *Sal*I; *Bgl*II (87C *D. melanogaster*); *Pst*I (87C *D. simulans*), and the 3' *Xba*I sites at 87A and 87C in *D. melanogaster* which we discuss in more detail later]. We have estimated the probability of a restriction site changing independently in both members of a pair of z elements to be 0.0297 (see Fig. 3 legend). The probability of making such an observation six times or more in 36 comparisons is 0.00062. Clearly, the loss of restriction sites in both z elements at a locus cannot be independent and we believe that a mechanism is operating which corrects sequence differences that arise between members of a z pair. Three intra-locus restriction site differences occur at the 87C1 locus of *D. mauritiana*. Presumably these have arisen since the last conversion event.

Observations have been made on vertebrate globins which have led to similar conclusions. Duplicated human α -globin

sequences are considerably more similar than expected from the observed species divergence^{25,26}. These sequences are arranged in tandem, and it has been suggested that repeated rounds of unequal crossing-over could explain the observed sequence similarity. This is supported by the discovery of rare variants with three α -globin sequences in tandem²⁵. However, unequal crossing-over cannot act on inverted repeat sequences. We consider a different mechanism to explain the sequence homogeneity we have observed between z elements within hsp 70 loci.

Figure 4 illustrates our proposed mechanism, which is based on previous models, to explain gene conversion during meiosis^{27,28}. Occasional folding back of a chromosome would bring the two z elements in a locus close together in the same orientation. Strand exchange occurring between proximal and distal genes combined with base-mismatch repair could lead to a locus containing a pair of similar z elements. Although Fig. 4 shows symmetric strand displacement, asymmetric exchange will also lead to sequence transfer between proximal and distal genes. The extent of gene conversion will depend on the length of branch migration so that it need not lead to exact similarity between sequence pairs. Note that unequal crossing-over cannot explain such incomplete exchange. A possible consequence of this model is the inversion of the DNA region between the two genes if the two sets of strand breakages occur on opposite strands (Fig. 4C, D). A similar mechanism could also operate between tandem genes except that symmetric exchange would give rise to deletions. Indeed, an intra-chromosomal recombination mechanism in *Drosophila* has been proposed to explain the reversion of certain tandem duplications at the *Bar* and *Beadex* loci in the absence of meiotic recombination and sister-strand exchange²⁹⁻³¹. We note that gene conversion has recently been invoked to explain discontinuities in sequence conservation found between human γ - and δ -globin genes³².

Presumably, the likelihood of gene conversion will depend on the distance between the two genes. Thus, one would expect the frequency of intra-locus exchange to be greatly diminished by the insertion of the large $\alpha\beta$ -containing region into the 87C1 locus of *D. melanogaster*. Sequence comparison between gene pairs at 87A7 would allow an estimate of the rate of exchange between two closely spaced hsp70 genes. If the insertion of the $\alpha\beta$ domain into the 87C1 locus causes a dramatic reduction of intra-locus conversion, the rate of nucleotide substitution between proximal and distal hsp70 genes will rise. Determination of the degree of sequence divergence between proximal and distal hsp70 genes at 87C1 in *D. melanogaster* will allow estimation of the time of insertion of the $\alpha\beta$ domain.

Sequence homogeneity between *z* elements from different loci

In a recent study³³, cloned *hsp70* coding sequences derived from *D. melanogaster* were hybridized to the polytene chromosomes of the unrelated species *Drosophila pseudoobscura*. This revealed two linked sites of sequence homology at heat-induced chromosome puffs. These are located on chromosome 2 in *D. pseudoobscura* which corresponds to the chromosome arm (3R) in *D. melanogaster* that contains the *hsp70* loci. This suggests that the duplication event which gave rise to two *hsp70* loci is very ancient, the *obscura* and *melanogaster* lineages probably diverging in the early Oligocene some 35×10^6 yr ago³⁴. Despite this long separation the *z* elements of the two loci are very similar.

Comparison of the *z* elements in the three species studied shows no greater divergence between the two *hsp70* loci than that between species at one locus (Fig. 3). In particular, only one restriction site differentiates *z* elements from the two loci of *D. simulans*. This suggests that a 'correction' mechanism is acting between the two *hsp70* loci, despite their being about 500 kilobases apart. Further evidence of such sequence exchange comes from the distribution of *Xba*I sites in the *z* elements.

All *z* elements in *D. melanogaster* have an *Xba*I site some 300 bases upstream from the terminal *Sal*I site^{8,11,35}. This site is absent from all copies in *D. simulans* and *D. mauritiana*. Three alternative origins of this distribution can be envisaged: (1) the 3'

*Xba*I site was present in the 'ancestral' *z* element and was lost from all copies in the lineage which led to *D. simulans* and *D. mauritiana*. (2) The loss of this site is the result of site-specific DNA modification rather than a coincidental change in the primary structure of the DNA of all *z* elements in *D. simulans* and *D. mauritiana*. (3) The 3' *Xba*I site arose in a *z* element at one locus in *D. melanogaster*, spread through that locus by an intra-locus conversion mechanism and then to the other *hsp70* locus by an inter-locus gene conversion event.

We have data which seem to rule out the first two possibilities. First, none of the slightly more distantly related species, *D. teisseri*, *D. yakuba* or *Drosophila erecta*, has a 3' *Xba*I site in its *z* elements (A.J.L.B., unpublished observations). According to the first scheme, one or more of these would also be likely to have retained this site, as they are no more closely related to *D. simulans* than to *D. melanogaster*¹⁴.

We have examined the possibility of base modification by analysing the *z* elements of *D. melanogaster*/*D. simulans* hybrid flies, and find that none of the sequences derived from *D. simulans* has a 3' *Xba*I site and all those from *D. melanogaster* are cut by *Xba*I (codominant inheritance). It is unlikely that the lack of a 3' *Xba*I site in species other than *D. melanogaster* is due to a DNA modification system, although we cannot completely exclude an enzyme active only on previously modified sites.

The third alternative remains: that gene conversion events occur between the two *hsp70* loci as well as within a locus. These must be rarer than intra-locus exchanges because we have found only one *hsp70* locus with dissimilar *hsp70* genes whereas the *z* elements differ between 87A and 87C1 in all three species studied. The relative rarity of inter-locus exchange reflects the lower likelihood of synapsis between non-adjacent genes. Although we cannot directly determine the frequency of gene conversion between the two *hsp70* loci, we can estimate the time of the most recent exchange. The non-protein-coding sequences in *z* at 87A7 and 87C1 have diverged by 15%^{10,20}. This greatly exceeds our estimate of the degree of nucleotide divergence in noncoding DNA between these species (Table 1), indicating that the last complete conversion between the sequences at 87A7 and 87C1 greatly predated the separation of *D. melanogaster* and *D. simulans*. The conversion of the 3' *Xba*I site in *D. melanogaster* that must have occurred after speciation illustrates that conversion events need only exchange parts of the *hsp70* gene. Recently, Scherer and Davis have demonstrated³⁶ gene conversion in yeast between unlinked *his3* genes associated with the yeast transposon, Ty1. Evidence for gene conversion events between unlinked suppressor genes in *Saccharomyces pombe* has also been reported³⁷.

The best characterized precedent for gene conversion between distant loci is the switching of mating type in yeast where conversion of the mating-type locus *MAT* by normally silent copies at *HML* and *HMR* has been demonstrated genetically and biochemically^{38,39}. Attempts to detect extra-chromosomal mating-type sequences were unsuccessful³⁸. The alternative model involves conventional recombinational mechanisms and resembles an inter-locus version of our model for intra-locus exchange^{40,41}. We believe that the mechanisms of intra- and inter-locus conversion of *hsp70* genes are similar and differ mainly in frequency. Our results clearly indicate that these frequencies are sufficient to be of considerable evolutionary significance.

We thank Drs W. Bender, D. Charlesworth, R. Holliday, C. Langley and J. Williams, and our colleagues in the laboratory for stimulating discussions and criticism of the manuscript. We also thank Drs M. Ashburner, J. P. W. Young and R. A. Voelker for fly stocks, Dr M. Meselson for plasmids pPW229.1 and pPW232.1, and Drs P. Schedl and A. Udvardy for plasmids 56H8 and cHS2S2.

Received 5 November 1980; accepted 23 January 1981.

1. Ashburner, M. & Bonner, J. J. *Cell* **17**, 241-254 (1979).
2. Livak, K., Freund, R., Schweber, M., Wensink, P. C. & Meselson, M. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5613-5617 (1978).

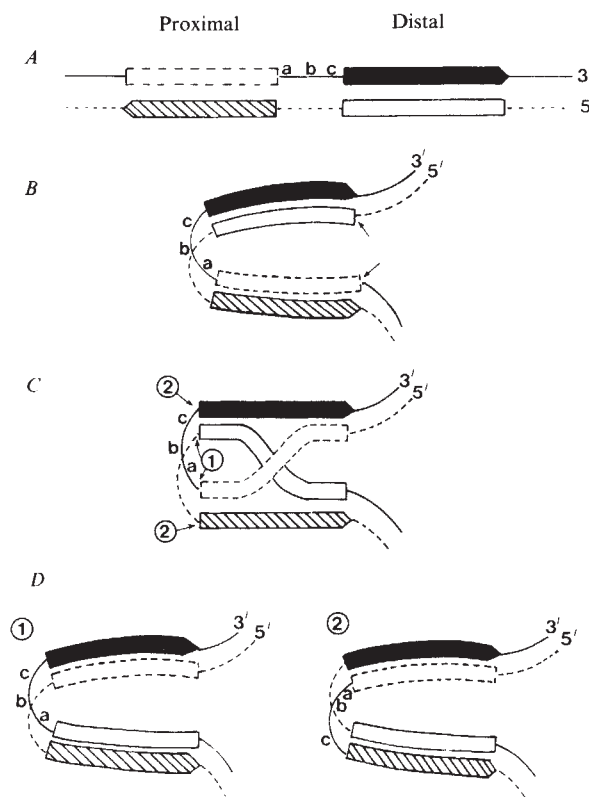


Fig. 4 Mechanism for intrachromosomal gene conversion. Filled boxes indicate coding sequences, transcribed in the direction of the arrow head, and empty ones the complementary sequences. Arrows indicate the sites of DNA strand breaks. The letters a, b and c indicate the orientation of the intergenic region. A, *hsp70* gene alignment; B, double-stranded hairpin loop formation; C, strand displacement and exchange with two possible secondary sites of strand breakage; D, the two products. DNA repair will correct the base mismatch at the proximal and distal genes in the same direction in a proportion of these events. Note the inversion of the intervening markers in D(2). The mechanism depicted would operate as readily between sister chromatids as between sequences on the same chromatid. Indeed, a version which converts between paired homologues during meiosis can be envisaged.

3. Artavanis-Tsakonas, S., Schedl, P., Mirault, M.-E., Moran, L. & Lis, J. *Cell* **17**, 9-18 (1979).
4. Craig, E. A., McCarthy, B. J. & Wadsworth, S. C. *Cell* **16**, 575-588 (1979).
5. Ish-Horowitz, D. *et al.* *Cell* **17**, 565-571 (1979).
6. Lewis, M., Helmsing, P. J. & Ashburner, M. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3604-3608 (1975).
7. Lis, J., Prestidge, L. & Hogness, D. S. *Cell* **14**, 901-919 (1978).
8. Ish-Horowitz, D., Pinchin, S. M., Schedl, P., Artavanis-Tsakonas, S. & Mirault, M.-E. *Cell* **18**, 1351-1358 (1979).
9. Mirault, M.-E., Goldschmidt-Clermont, M., Artavanis-Tsakonas, S. & Schedl, P. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5254-5258 (1979).
10. Holmgren, R., Livak, K., Morimoto, R., Freund, R. & Meselson, M. *Cell* **18**, 1359-1370 (1979).
11. Moran, L. *et al.* *Cell* **17**, 1-8 (1979).
12. Goldschmidt-Clermont, M. *Nucleic Acids Res.* **8**, 235-252 (1980).
13. Ish-Horowitz, D. & Pinchin, S. M. *J. molec. Biol.* **142**, 231-245 (1980).
14. Lemeunier, F. & Ashburner, M. *Proc. R. Soc. B* **193**, 275-294 (1976).
15. Southern, E. M. *J. molec. Biol.* **98**, 503-518 (1975).
16. Lis, J. T., Ish-Horowitz, D., Pinchin, S. M. & Hogness, D. S. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
17. Schedl, P. *et al.* *Cell* **14**, 921-929 (1978).
18. Martin, G. *et al.* (in preparation).
19. Gausz, J. *et al.* *Genetics* **93**, 917-934 (1979).
20. Karch, F. & Torok, I. (in preparation).
21. Carson, H. L. *Nature* **259**, 395-396 (1976).
22. Ohnishi, S., Leigh Brown, A. J., Voelker, R. A. & Langley, C. H. *Genetics* (submitted).
23. Kohn, D. E. *O. Rev. Biophys.* **3**, 327-375 (1970).
24. Benveniste, R. E. & Todaro, G. J. *Nature* **261**, 101-108 (1976).
25. Zimmer, E. A., Martin, S. L., Beverley, S. M., Kan, Y. W. & Wilson, A. C. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2158-2162 (1980).
26. Lauer, J., Shen, C.-K. J. & Maniatis, T. *Cell* **20**, 119-130 (1980).
27. Holliday, R. *Genet. Res.* **5**, 282-304 (1964).
28. Meselson, M. S. & Radding, C. M. *Proc. natn. Acad. Sci. U.S.A.* **72**, 358-361 (1975).
29. Peterson, H. M. & Laughnan, J. R. *Proc. natn. Acad. Sci. U.S.A.* **50**, 126-133 (1963).
30. Laughnan, J. R. & Gabay, Molec. gen. *Genet.* **108**, 93-96 (1970).
31. Green, M. M. *Molec. gen. Genet.* **103**, 209-217 (1968).
32. Slightom, J. L., Blechl, A. E. & Smithies, O. *Cell* **21**, 627-638 (1980).
33. Pierce, D. A. & Lucchesi, J. C. *Chromosoma* **76**, 245-254 (1980).
34. Throckmorton, L. H. in *Handbook of Genetics* Vol. 3 (ed King, R. C.), 421-470 (Plenum, New York, 1974).
35. Torok, I. & Karch, F. *Nucleic Acids Res.* **8**, 3105-3123 (1980).
36. Scherer, S. & Davis, R. W. *Science* **209**, 1380-1384 (1980).
37. Hofer, F. *et al.* *Curr. Genet.* **1**, 45 (1979).
38. Hicks, J., Strathern, J. & Klar, A. J. S. *Nature* **282**, 478-483 (1979).
39. Nasmyth, K. A. & Tatchell, K. *Cell* **19**, 753-764 (1980).
40. Haber, J. E., Rogers, D. T. & McCusker, J. H. *Cell* **22**, 277-289 (1980).
41. Klar, A. J. S., McIndoo, J., Strathern, J. N. & Hicks, J. B. *Cell* **22**, 291-298 (1980).
42. Nei, M. & Li, W.-H. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5269-5273 (1979).
43. Wahl, G. M., Stern, M. & Stark, G. R. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3683-3687 (1979).

LETTERS

Detection of low order p-modes in brightness fluctuations of the Sun

Franz-Ludwig Deubner

Institut für Astronomie und Astrophysik, 8700 Würzburg, FRG

The discovery of radial and non-radial global oscillations of the Sun through various types of detector systems has opened the field of solar seismology. Of particular importance for studies of the deep interior of stars are the detection and analysis of low harmonic order pulsations such as the low order p-modes recently found on the Sun^{1,2} by line shift measurements. These pulsations have now also been detected in the solar brightness fluctuations. Through the reflected light of Uranus and Neptune the Sun was observed as though it were a distant star. The positive results suggest that broad-band photometry has the greatest potential for developing seismometry of main sequence stars.

High overtone pulsations of the Sun, both radial and non-radial (but with low degree l), discovered through line shift observations of the integrated solar disk by means of a potassium resonance spectrometer, have been reported by Claverie *et al.*¹ and confirmed with similar instrumentation operated at the South Pole station by Fossat² and his colleagues (also see ref. 3 for review). The excellent solar results available suggest that this type of measurement should be extended to main sequence stars and other 'nonvariables'. However, because of the lack of photons, the sensitivity of the kind of equipment necessary to select a narrow-band spectral region for line shift analysis is barely sufficient to resolve motions with anticipated amplitudes smaller than a few metres per second, even for the brightest stars⁴. The integrated power in the low order p-mode band has an amplitude of $\sim 2 \text{ m s}^{-1}$. However, decomposition into its individual modes requires a sensitivity better than 20 cm s^{-1} , as in the solar case. For this reason, and also because of the more common availability of good broad-band photometric devices, it is of interest to study the possibility of detecting comparatively short period fluctuations of the integrated intensity related to the p-mode oscillations.

Clearly, one does not hope to see the geometrical effect of a change of the radius by $\delta r/r \approx 1.5 \times 10^{-8}$. It is the enhancement by a factor of 1,000-2,000 of the temperature fluctuation compared with the fractional change of the radius^{3,5} which brings the expected brightness fluctuations into the range of $\delta I/I \approx 2 \times 10^{-5}$ for vertical photospheric motions with the

amplitude (20 cm s^{-1}) and period (5 min) of the p-modes. Compared with common scintillation amplitudes of the order of 10^{-3} mag the expected signals still seem to be much too small to be detected.

For a successful observation of non-radial stellar p-modes it is important to resolve and identify the individual modes, in addition to detecting the signal. To obtain a resolution in frequency sufficient to resolve the solar p-mode spectrum (that is, better than $30 \mu\text{Hz}$), observing sessions of at least 10 h, preferably without interruption, are required. The duration of this observation will then also help to suppress the noise level due to atmospheric scintillation.

A recent search for intensity fluctuations of solar origin in daylight observations of the integral sunlight⁶ was unsuccessful. I have previously obtained⁷ data amenable to similar analysis at the ESO station in Chile, to investigate the presence of fluctuations with periods longer than 10 min in the solar luminosity, as monitored by broad-band dual beam photometry of the brightness of Uranus and Neptune using a 50-cm telescope. Combining observations of 10 successive nights with a mean observing duration of 7.8 produced a r.m.s. noise level of 6×10^{-5} in these data. In the present investigation the spectral analysis was repeated applying a least squares sine wave fitting technique to the complete set of data, to avoid the awkward consequences of the temporal window in a straight Fourier transform with zeroes padded in the data gaps. Power estimates were computed at $(1 \text{ cycle}/14,400) \text{ min}^{-1}$ intervals. The power spectrum is still very noisy and gives only a vague impression of regularly spaced features. Fortunately, asymptotically the p-mode eigenfrequencies are spaced at regular intervals, and the size of this interval is well enough established⁸ to permit a 'superposed frequency' analysis analogous to the superposed epoch technique used to detect non-sinusoidal periodic fluctuations. A superposed frequency analysis applied to a velocity power spectrum, which is analogous to the procedure here, was recently published by Grec *et al.*².

The result of this procedure, averaging 15 frequency segments Δf of $135.42 \mu\text{Hz}$ each, covering the range 2.083-4.167 mHz, is shown in Fig. 1b, accompanied by the 'unfolded' spectral data on the same frequency scale in Fig. 1a. At one position the signal is well above the 3σ level of this mean power distribution, at two more positions the signal exceeds the 2σ level. All three features are accompanied by ghosts at 10 frequency intervals' distance due to the periodic structure of the window function. (The apparent asymmetry of the ghost pattern can be shown by numerical simulation to be due partly to the shape of the window transform and partly to the comparatively high noise level of the data.) The principal frequencies falling next to the mean period of 5 min are 3.301, 3.309 and 3.377 mHz.

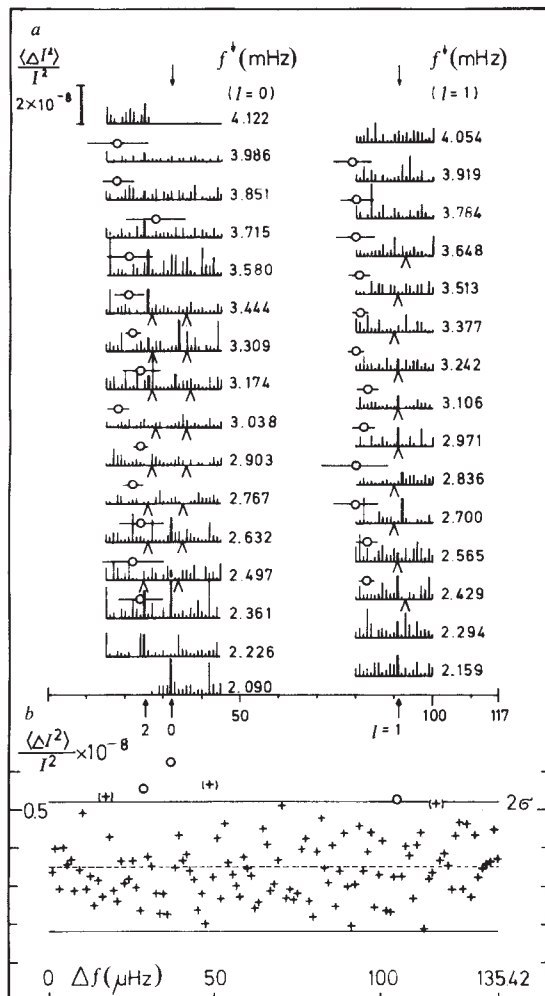


Fig. 1 Frequency spectrum of brightness fluctuations of the Sun. Spectral segments of length 135.42 μHz , shown in *a* are averaged to obtain the 'superposed' spectrum in *b*. Three peaks above the 2σ level are believed to represent sets of low order solar p-modes (ghost lines are bracketed). In *a*, the observed spectral frequencies which are believed to represent global p-modes are drawn with heavier lines. The frequency values given refer to the position within each segment marked by the arrows on top. The open dots with error bars in *a* are frequencies observed through line shift observations by Claverie *et al.*⁸. The arrow heads mark three sets of frequencies determined by Fossat (personal communication) from a 5 days continuous observation obtained at the South Pole. The numbered scale below the segmented spectrum indicates the 117 frequency intervals of $(1/14,400) \text{ min}^{-1}$ congruent with the folding interval of 135.42 μHz .

With the improved frequency resolution of this observation, one of the two sets of eigenfrequencies reported by Claverie *et al.*¹ is resolved into doublets at an interval of $\sim 8 \mu\text{Hz}$, similar to Fossat's² result of 10 μHz . Identification of these doublets with $l=0$ and $l=2$ modes is consistent with theoretical spectra published by Iben and Mahaffy⁹ and Christensen-Dalsgaard¹⁰. In detail, our observed frequencies agree quite satisfactorily with those obtained by Fossat (personal communication), but are systematically higher by 5–10 μHz than the values given by Claverie *et al.*⁸. To some extent, the accuracy of these latter data may suffer from insufficient spectral resolution. By standard statistical criteria, the significance of the three peaks emerging from the 2σ background of our superposed frequency analysis seems very marginal. We note, however, that of these peaks, two (with $l=1$ and 2) coincide within $\pm 2.25 \mu\text{Hz}$ with the most accurately observed superposed frequencies available. The probability of such a coincidence occurring by chance is 1.7×10^{-2} . Considering the statistical fluctuations of the amplitude estimates of a 10 days' data sample the mode identifications

given in Fig. 1 are consistent with the theoretical result that the mode numbers $l=0, 1$ and 2 are those most favoured by the spatial filter function both of full disk velocity and intensity observations¹¹. Finally, a phase difference of 3.5 yr in the solar activity cycle existing between measurements of intensity⁷ and velocity (E. Fossat, personal communication) should be considered when minor systematic frequency differences among these observations are confirmed.

After subtraction of the noise the observed mean power at the p-mode frequencies corresponds to a r.m.s. amplitude of $4\text{--}5 \times 10^{-5}$, in good agreement with our expectation. The improved signal-to-noise ratio obtained with stellar objects brighter than Uranus and Neptune, and with larger telescopes should facilitate the analysis by making the periodic structure stand out against the straight unfolded spectrum.

I thank G. Isaak and E. Fossat for unpublished results, and for exciting discussions.

Received 22 December 1980; accepted 12 February 1981.

1. Claverie, A., Isaak, G. R., McLeod, C. P., van der Raay, H. B. & Roca Cortes, T. *Nature* **282**, 591 (1979).
2. Grec, G., Fossat, E. & Pomerantz, M. *Nature* **288**, 541–544 (1980).
3. Gough, D. *Lecture Notes Phys.* **125**, 273 (1980).
4. Traub, W. A., Mariska, J. T. & Carleton, N. P. *Astrophys. J.* **223**, 583 (1978).
5. Hill, H. A., Rosenwald, R. D. & Caudell, T. P. *Astrophys. J.* **225**, 304 (1978).
6. Claverie, A., Isaak, G. R. & McLeod, C. P. *Proc. 14th ESLAB Symp.*, Scheveningen (in the press).
7. Deubner, F. -L. *Astr. Astrophys.* **57**, 317 (1977).
8. Claverie, A., Isaak, G. R., McLeod, C. P., van der Raay, H. B. & Roca Cortes, T. *Proc. 14th ESLAB Symp.*, Scheveningen (1980).
9. Iben, I. Jr & Mahaffy, J. *Astrophys. J.* **209**, L39 (1976).
10. Christensen-Dalsgaard, J. & Gough, D. O. *Nature* **288**, 544 (1980).
11. Hill, H. A. in *The New Solar Physics* (ed. Eddy, J. A.) 135 (Westview Press, Colorado, 1978).

Solar neutrinos and turbulent diffusion

E. Schatzman

Observatoire de Nice, BP 252, Nice Cedex 06007, France

A. Maeder

Observatoire de Genève, 1290 Sauverny, Switzerland

Stellar rotation generates instabilities¹ which have not been taken into account in classical computations of stellar evolution. In particular, these instabilities may produce turbulent diffusion with important consequences over long time scales. Thus it has been shown^{2,3} that the loss of angular momentum which occurs in the external stellar layers due to stellar wind and magnetic braking⁴ leads to a redistribution of angular momentum. We show here that there are observational and theoretical reasons to suspect that a slow diffusion process due to a mild turbulence is operating in stellar interiors. Turbulent diffusion may have important consequences for the solar neutrinos problem and the (³He/⁴He) ratio at the solar surface.

The necessary and sufficient condition for the generation of turbulence has been often discussed. The presence of shear is a necessary but not sufficient condition; the Richardson criterion⁵ has also to be satisfied. But, when account^{6–8} is taken of the heat exchange between the moving fluid and the surrounding medium the Richardson criterion has to be replaced by the less stringent Townsend criterion or flux Richardson criterion. Eliassen and Kleinschmidt⁹ used the Townsend criterion to describe the maintenance of turbulence which is the same as that used by Zahn⁸. These conditions naturally differ from those of Layzer¹⁰, who has assumed that no heat exchange takes place during the motion of the fluid, a more stringent requirement.

The Rayleigh condition, states that, as long as $\delta\Omega/\delta z = 0$, the fluid is stable provided $\delta(\Omega\varpi^2)/\delta\varpi < 0$, Ω being the angular velocity and ϖ the distance to the axis. This last condition is violated if the star, like the Sun, is losing angular momentum. But as soon as $\delta\Omega/\delta z \neq 0$ the rotational motion is dynamically unstable¹¹. It has been shown that there is no meridional circulation which can be stable everywhere in a star^{12,13}.

We show here that the absence of turbulence in the radiatively stable region of the Sun would be remarkable¹⁴. We assume that the instabilities associated with the shear flows in differentially rotating stars generate a mild turbulence, the effects of which may be treated¹⁵ as a diffusion process in stellar evolution.

Several astrophysical observations seem to support the mechanism of turbulent diffusion. Schatzman¹⁶ suggested that turbulence could inhibit gravitational (and radiative) sorting of the elements, and Baglin¹⁷ showed that meridional circulation for fast-rotating stars can induce a sufficiently high level of shear flow turbulence to avoid the sorting of elements. The Li deficiency in giants of spectral types F and G has been measured by Alschuler¹⁸. Assuming that these giants belong to the first ascending branch, and that their mass can be derived from Iben's¹⁹ evolutionary tracks, the Li deficiency can be explained by Li burning during the main sequence phase: Li is carried by turbulent diffusion from the outer layers of the star towards the region where it is nuclearly processed and destroyed. For masses $M \leq 1.6 M_{\odot}$, the empirical estimates of this rate of Li destruction on the main sequence²⁰ essentially agree with the theoretical values based on the turbulent diffusion model.

The ^{13}C excess in giants of the first ascending branch²¹, can be accounted for by turbulent diffusion²² outwards from the region where ^{12}C is nuclearly processed and partly transformed into ^{13}C .

Note that the equations for a slow diffusion process in stellar evolution are quite different from those for convective mixing. In particular, homogenization in finite volumes²³ does not properly mimic a slow diffusion process. This can be explained by a simple physical argument. The complete mixing over a fraction q_m of the solar mass²³⁻²⁶ generates a step function of the chemical composition, with a jump in the chemical composition at the mass fraction q_m . For solar models, with the constraint that the luminosity is $L = L_{\odot}$ for $t = 4,650$ Myr, we have $L \sim X_{1\text{eff}}^2 T_c^4$, where $X_{1\text{eff}}$ is some convenient average value of the hydrogen concentration X_1 . This means that $T_c \sim X_{1\text{eff}}^{-1/2}$. With the power law approximation²⁷ $\phi_{\nu} \sim T_c^{20} X_4 (1 + X_1)^{-1}$, the reduction in the neutrino flux is given by

$$\frac{(\phi_{\nu})_{\text{mixed}}}{\phi_{\nu}} \approx \left(\frac{X_{1c}}{X_{1\text{eff}}} \right)^{-10} \frac{(X_4)_{\text{mixed}}}{X_4} \frac{(1 + X_1)}{(1 + X_1)_{\text{mixed}}}$$

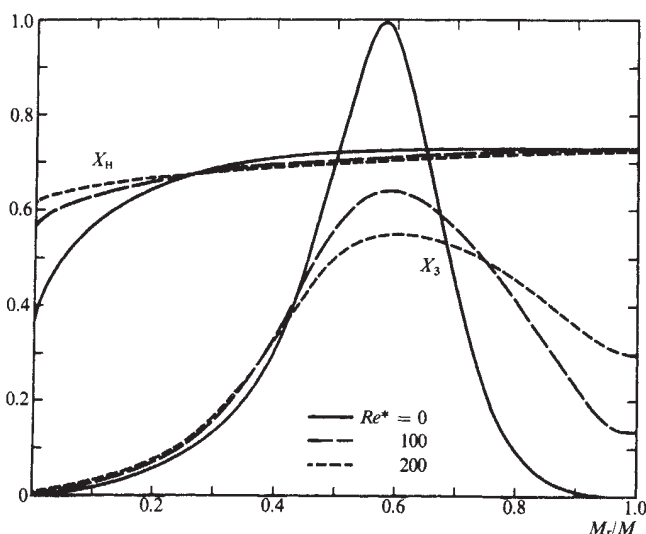


Fig. 1 Curves X_H give the hydrogen concentration as a function of the mass fraction. Turbulent diffusion brings hydrogen in the core of the Sun. Curves X_3 give the ^3He concentration as a function of the mass fraction. Turbulent diffusion smoothes the ^3He peaks, with ^3He appearing at the surface and the concentration of ^3He increasing by an appreciable amount in the centre. The maximum X_3 corresponds to 2.8×10^{-3} . The increase of the concentration of the nuclear fuel in the centre is balanced by a decrease of the central temperature which produces a drastic reduction in the neutrino flux. Meanwhile, ^3He appears at the surface of the Sun.

The ratio $(X_{1c}/X_{1\text{eff}})$ is always smaller for fully mixed models to $M = q_m M_{\odot}$ than for models where diffusion mixing has taken place. This is because the step function of the chemical composition is not a true description of the distribution of the chemical composition due to the diffusion process.

Schatzman¹⁵ and Genova and Schatzman²² demonstrated that a turbulent diffusion coefficient $D = Re^* \nu$, can be used where ν is the microscopic viscosity (molecular plus radiative) and Re^* is a critical Reynolds number of the order of 100. The changes of compositions in stellar models due to both the nuclear reactions and the turbulent transport may be expressed as:

$$\rho \frac{\partial X_1}{\partial t} = \frac{1}{r^2} \frac{\partial}{\partial r} r^2 \rho D \frac{\partial X_1}{\partial r} + \rho A_1 m_H (-3R_{11} + 2R_{33} - R_{34})$$

$$\rho \frac{\partial X_3}{\partial t} = \frac{1}{r^2} \frac{\partial}{\partial r} r^2 \rho D \frac{\partial X_3}{\partial r} + \rho A_3 m_H (R_{11} - 2R_{33} - R_{34})$$

$$\rho \frac{\partial X_4}{\partial t} = \frac{1}{r^2} \frac{\partial}{\partial r} r^2 \rho D \frac{\partial X_4}{\partial r} + \rho A_4 m_H (R_{33} + R_{34})$$

X_1 , X_3 and X_4 are functions of r and t and represent the abundances, expressed in mass fraction, of ^1H , ^3He and ^4He respectively. ρ is the density and the R_{ij} terms are the nuclear reaction rates²⁸ for the species ij . An efficient matrix scheme for the diffusion equations²⁹ has been incorporated into our stellar evolution code.

Models with $Re^* = 200$, 100 and 0 (standard case) for an initial composition $X_1 = 0.73$, $X_4 = 0.25$ and $Z = 0.02$ have been computed. We note the following main results. As shown in Fig. 1, which gives the distribution of X_1 and X_3 in the Sun at $t = 4,600$ Myr, there is much more central hydrogen in the present Sun when diffusion is included. While the central X_1 is only 0.359 in the standard case, it is 0.569 with $Re^* = 100$. This excess of central X_1 evidently results¹⁵ from the inward transport due to a positive abundance gradient. Outside the central regions, the differences are smaller and of opposite sign. Overall, turbulent diffusion makes the X_1 distribution much flatter.

The internal distribution of ^3He shows a peaked distribution in the standard model (see Fig. 1). With turbulent diffusion, with $Re^* = 100$ and 200, the distribution of ^3He below $M_r/M = 0.4$ is still mainly determined by the rapid equilibrium between creation and destruction as in the standard case. Above $M_r/M = 0.4$ ($T < 8.3 \times 10^6$ K), the transport by turbulent diffusion tends to unify the ^3He distribution³⁰: the peak is reduced and ^3He is spread throughout the star. Figure 1 shows that the surface ratio $^3\text{He}/^4\text{He}$ is considerably increased by turbulent diffusion. An acceptable observational value of this ratio^{31,32} is $^3\text{He}/^4\text{He} \approx 4 \times 10^{-4}$. This value is $\sim 2-4$ times larger³³ than that of the protosolar material and implies a surface enrichment in ^3He during solar evolution. This result supports the occurrence of turbulent diffusion with a value of $Re^* \approx 60$.

The decrease of the mean molecular weight in the interior due to diffusion leads to a decrease of the central temperature and density and, in models with turbulent diffusion, the central concentration of ^3He is increased. Some characteristics of the models at $t = 4,600$ Myr are given in Table 1. We see that the neutrinos flux is smaller by a factor of 8.1 for $Re^* = 200$, 4.9 for $Re^* = 100$ and 2.6 for $Re^* = 50$ relative to the flux obtained in the standard case $Re^* = 0$.

The constraint on the solar luminosity usually fixes the helium content in the initial model; a change of X_4 by -0.01 reduces the luminosity by $\log L = -0.028$. Table 1 shows that diffusion models may fulfill the requirement on solar luminosity with a helium content differing only by 0.01–0.02 from that of the standard case.

Estimates have been made of the possibility of choosing a Z (heavy elements), X_4 (Helium) and Re^* to fulfill the three main requirements: the value of the solar luminosity, the ^3He concentration at the surface and the solar neutrino flux. As a whole, models with turbulent diffusion with Re^* in the range

Table 1 Properties of the standard model of the Sun and of the models with turbulent diffusion mixing at $t = 4,600$ Myr

Re^*	$\log T_c$	$(X_1)_c$	$(X_3)_c$	N_ν (SNU)	$(X_3/X_4)_s$	$\log L/L_\odot$
0	7.1944	0.3585	$8 \cdot E-6$	11.66	—	0.048
100	7.1681	0.5685	$2.1 \cdot E-5$	2.38	$1.57 \cdot E-3$	0.009
200	7.1601	0.6164	$2.6 \cdot E-5$	1.43	$3.38 \cdot E-3$	0.003

$$X = 0.73, Y = 0.25, Z = 0.02.$$

$(X_1)_c$ is the hydrogen concentration in the centre; $(X_3)_c$ is the ^3He concentration in the centre; $(X_3/X_4)_s$ is the surface concentration ratio of ^3He and ^4He at the surface.

50–100 seemed to fulfill the main solar constraints for a standard composition.

Our solution overcomes the difficulties of the low X_4 and low Z solar models³⁴. However, recent observations of the solar oscillations³⁵ have led Christensen-Dalsgaard and Gough³⁶ to describe solar models which agree with standard models. The difficulty³⁷ due to migration of heavy elements through a shallow convective zone can be avoided. The same model explains the depletion of lithium at the surface of the Sun.

From dimensional arguments alone we can see that the depletion of lithium would be of the order of $\exp[-(t_\odot/\tau_{\text{Li}})]$ and the depletion of Be would be of the order of $\exp[-(t_\odot/\tau_{\text{Be}})]$. τ_{Li} and τ_{Be} are of the order of $(\Delta r)_{\text{Li}}^2 D_{\text{turb}}^{-1}$ and $(\Delta r)_{\text{Be}}^2 D_{\text{turb}}^{-1}$ respectively, where $(\Delta r)_{\text{Li}}$ and $(\Delta r)_{\text{Be}}$ are the respective distances from the bottom of the hydrogen convective zone to the Li burning and Be burning regions. From the values of the Li depletion factor ($\approx 4.8 \times 10^2$) and of the Be depletion factor (≈ 2), and knowing the temperatures of the Li and Be burning regions ($T_{\text{Li}} = 2.4 \times 10^6$ K; $T_{\text{Be}} = 3.5 \times 10^6$ K), we can derive both where the hydrogen convective zone ends and the value of the turbulent diffusion coefficient (in fact, we do not obtain the bottom limit of the convective zone, but the point at which the penetrative convection T_{PCE} becomes inefficient):

$$T_{\text{PCE}} = 2.19 \times 10^6 \text{ K} \quad R - r_{\text{PCE}} = 2.06 \times 10^5 \text{ km}$$

$$T_{\text{Li}} = 2.4 \times 10^6 \text{ K} \quad r_{\text{PCE}} - r_{\text{Li}} = 0.2 \times 10^5 \text{ km}$$

$$T_{\text{Be}} = 3.5 \times 10^6 \text{ K} \quad r_{\text{PCE}} - r_{\text{Be}} = 1.24 \times 10^5 \text{ km}$$

These values are compatible with those of Christensen-Dalsgaard and Gough³⁶. From the value of D_{turb} we obtain $Re^* \approx 85$ which is compatible with the values derived from the other solar constraints.

Meridional circulation³⁸ does not yield the same result. Meridional circulation drives the matter down from the bottom of the hydrogen convective zone to where the Li and Be are burning, and will not differentiate between Li and Be: the characteristic time of depletion being determined by the velocity of the circulation would be the same for both elements. Unless there was a very peculiar pattern of the meridional circulation, we can consider that this is a very strong argument in favour of turbulent diffusion and against meridional circulation²⁸.

We now comment on the role of turbulent diffusion in relation to the neutrino mass³⁹. Proposed values of the neutrino mass favour the suggestion^{40–42} that the neutrino oscillates between the Sun and the Earth. However, the non-zero mass of the neutrino does not necessarily imply an oscillation⁴³. It is therefore worth considering the consequence of change in the stellar models by introducing a simple physical effect and studying its implication for the Sun and the stars. It is difficult⁴⁴ to resolve the difference between predictions based on the solar models and observations by invoking neutrino oscillations only, if there are only three kinds of neutrinos coupled to each other. The probability of finding an electron neutrino after a time t is estimated by Bahcall and Frautschi⁴⁵ to be 0.5 and by Barger *et al.*⁴⁶ to be 0.62. With the latter value, this decreases for a standard model⁴⁴ the observable neutrino flux to 4.65 SNU, far above the observed flux.

How will the Sun evolve within the above framework? This question is particularly important in relation to stellar observations. Clearly, if turbulent diffusion were always maintained, a $1 M_\odot$ star would never become a red giant as the evolutionary time scale for such a star is larger than the diffusion time scale. The star would evolve towards the region of blue stragglers. To obtain an evolution towards the giant branch, turbulent diffusion has to be stopped. The progressively increasing μ -gradient during evolution probably⁸ has a stabilizing effect on the shear flow and consequently the turbulent diffusion would cease when the μ -gradient became large enough.

This suggests the existence of two groups of stars: (1) those with a very low differential rotation, which pursue a standard evolution; (2) those with a high degree of differential rotation, which evolve towards the blue-straggler region. There might also exist a wide range of intermediate cases for which the evolution towards the red-giant branch is only delayed and not stopped completely. This possibility seems very likely because cluster observations indicate⁴⁷ that the width of the main-sequence band is considerably larger than that given by standard models.

Although the evolution of the Sun is not very advanced, previous considerations suggest that this evolution would not necessarily follow route (1) above. In fact, we cannot yet estimate properly the period of rotation of the solar core. The minimum requirement for maintaining the turbulence is that the flux Richardson criterion is satisfied. This somehow gives the minimum angular velocity which might be present inside the Sun. Following Zahn⁸ this can be written

$$\sigma \frac{d\Omega}{d\tau} \geq (\sigma Re_{\text{crit}}) \left(\frac{g}{H_P} \right) (\nabla_{\text{ad}} - \nabla_{\text{rad}})$$

where $\sigma = (\nu/\lambda_{\text{th}})$, ratio of viscosity to the thermal diffusivity is the Prandtl number, Re_{crit} the critical Reynolds number, H_P the pressure scale height and $\nabla = (d \log T / d \log P)$. This leads to a minimum value of Ω in the solar core, corresponding to a period of 2–3 days, in agreement with a vanishingly small oblateness. The turbulent diffusion coefficient for angular momentum therefore does not have a simple relationship with the turbulent diffusion coefficient for chemical species. Some feedback must exist, between loss of angular momentum, transport of angular momentum and turbulence which presumably stabilizes the critical Reynolds number Re^* around 100.

Received 3 November 1980; accepted 8 February 1981.

1. Tassoul, J. L. *Theory of Rotating Stars* (Princeton University Press, 1978).
2. Howard, L. N., Moore, D. W. & Spiegel, E. A. *Nature* **214**, 1297–1299 (1967).
3. Bretherton, F. P. & Spiegel, E. A. *Astrophys. J. Lett.* **153**, 177–8 (1978).
4. Schatzman, E. *Ann. Astrophys.* **25**, 18–29 (1962).
5. Richardson, L. F. *Proc. R. Soc. A* **97**, 354 (1920).
6. Townsend, A. A. *J. Fluid Mech.* **4**, 361 (1958).
7. Moore, D. W. & Spiegel, E. A. *Astrophys. J.* **139**, 48 (1964).
8. Zahn, J. P. *IAU Symp.* No 59, 185 (1974).
9. Eliassen, A. & Kleinschmidt, E. *Hdl d. Physik*, **48**, 45 (1957).
10. Layzer, D. in *Ionospheric Sporadic E* Vol. 2, 258 (Pergamon, Oxford, 1962).
11. Goldreich, P. & Schubert, G. *Astrophys. J.* **150**, 571 (1967).
12. Randers, G. *Astrophys. J.* **95**, 454 (1942).
13. Tassoul, J. L. *Theory of Rotating Stars* (Princeton University Press, 1978).
14. Kippenhahn, R. *IAU Symp.* No 66, 20 (1974).
15. Schatzman, E. *Astr. Astrophys.* **56**, 211–218 (1977).
16. Schatzman, E. *Astr. Astrophys.* **3**, 331–346 (1969).
17. Baglin, A. *Astr. Astrophys.* **19**, 45–50 (1972).
18. Alschuler, W. R. *Astrophys. J.* **195**, 649–660 (1975).
19. Iben, I. *Astrophys. J.* **142**, 1447 (1965); **147**, 650 (1967).
20. Scalo, J. M. & Miller, G. E. *Astrophys. J.* **239**, 953–960 (1980).
21. Lambert, D. L. *Mem. Soc. R. Sci. Liège* **9**, 405–415 (1976).
22. Genova, F. & Schatzman, E. *Astr. Astrophys.* **78**, 323–327 (1979).
23. Shaviv, G. & Salpeter, E. E. *Phys. Rev. L.* **21**, 1602–1605 (1968).
24. Ezer, D. & Cameron, A. G. W. *Astrophys. Lett.* **1**, 177 (1968).
25. Shaviv, G. & Beaudet, G. *Astrophys. Lett.* **2**, 17 (1968).
26. Bahcall, J. N., Bahcall, N. A. & Ulrich, R. K. *Astrophys. Lett.* **2**, 91 (1968).
27. Iben, I. Jr *Ann. Phys.* **54**, 164 (1969).
28. Fowler, W. A., Caughlan, G. R. & Zimmerman, B. A. *A. Rev. Astr. Astrophys.* **13**, 69 (1975).
29. Angrand, F. thesis, Univ. Pierre et Marie Curie, Paris (1980).
30. Schatzman, E. *Astrophys. Lett.* **3**, 139–140 (1969).
31. Geiss, J., Buhler, F., Cerutti, H., Eberhardt, P. & Fillex, Ch. *Apollo 16 Prel. Sci. Rep.* (NASA SP-315, 1972).
32. Hall, D. N. B. *Astrophys. J.* **197**, 509 (1975).
33. Geiss, J. & Bochsler, P. Preprint (1980).
34. Christensen-Dalsgaard, J., Gough, D. O. & Morgan, J. G. *Astr. Astrophys.* **73**, 121–128 (1979).

35. Grec, G., Fossat, E. & Pomerantz, M. *Nature* **288**, 541 (1980).
36. Christensen-Dalsgaard, J. & Gough, D. O. *Nature* **288**, 544 (1980).
37. Michaud, G. *Nature* **266**, 433 (1977).
38. Vauclair, S., Vauclair, G., Schatzman, E. & Michaud, G. *Astrophys. J.* **223**, 567–582 (1978).
39. Close, F. E. *Nature* **284**, 507–508 (1980).
40. Maki, Z., Nakagawa, M. & Sakato, S. *Prog. theor. Phys.* **28**, 870 (1962).
41. Pontecorvo, B. *Z. exp. theor. Phys.* **53**, 1717 (1967); *Soviet Phys. JETP* **26**, 984 (1968).
42. Gribov, V. & Pontecorvo, P. *Phys. Lett.* **28B**, 493–496 (1969).
43. Mohapatra, R. W. & Senjenovic, G. *Phys. Rev. Lett.* **44**, 912–916 (1980).
44. Bahcall, J. N. *et al.* *Phys. Rev. Lett.* **45**, 945 (1980).
45. Bahcall, J. N. & Frautschi, S. C. *Phys. Lett.* **29B**, 623 (1969).
46. Barger, V., Whisnaut, K. & Phillips, R. J. N. *Phys. Rev. D* **22**, 1636 (1980).
47. Maeder, A. *Astr. Astrophys.* **32**, 177 (1974).

Solvent inclusion and disorder effects on an isostructural series of organic conductors

G. J. Ashwell

Department of Chemistry, Sheffield City Polytechnic, Pond Street, Sheffield S1 1WB, UK.

Interest in quasi-one-dimensional organic conductors has been renewed by the discovery¹ of superconductivity in (tetramethyltetraselenafulvalene)₂PF₆, (TMTSF)₂PF₆. At ambient pressure (TMTSF)₂PF₆ undergoes a Peierls distortion at 18 K (ref. 2) whereas under modest pressure the properties remain metallic with the onset of superconductivity at 0.9 K (ref. 1). There is now an incentive to synthesize other organic metals in which the Peierls distortion is suppressed in the hope that these may also superconduct under pressure. So far (DEPE)(TCNQ)₄(H₂O)_x is the only example^{3,4} to show a monotonic increase in conductivity, characteristic of a metal, on cooling at ambient pressure to below 30 mK. This metallic behaviour has, however, been disputed⁵. I verify here the unique temperature dependence of conductivity of (DEPE)(TCNQ)₄(H₂O)_x and report the electrical and structural properties of the isostructural congeners (DMPP)(TCNQ)₄(H₂O)_x and (DRPA)(TCNQ)₄(H₂O)_x where R=M (methyl) and E (ethyl) (see Table 1 for full names).

Needle-shaped crystals, 2 × 0.1 × 0.01 mm, were prepared by a method reported previously⁴ for bipyridinium-TCNQ salts. The crystals were fibrous and frequently bowed indicating a weak coupling between chains. The stoichiometries were

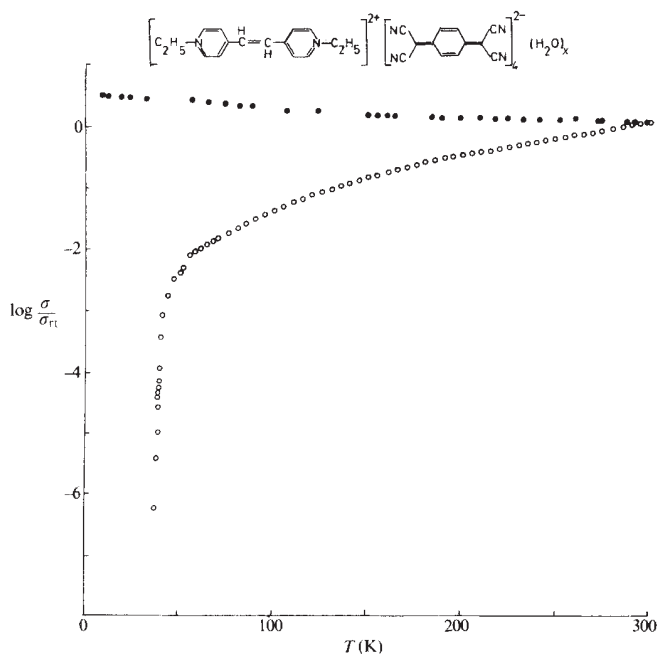


Fig. 1 Variation of the normalized conductivity with temperature of two crystals of (DEPE)(TCNQ)₄(H₂O)_x, one semiconducting (○) and the other metallic (●).

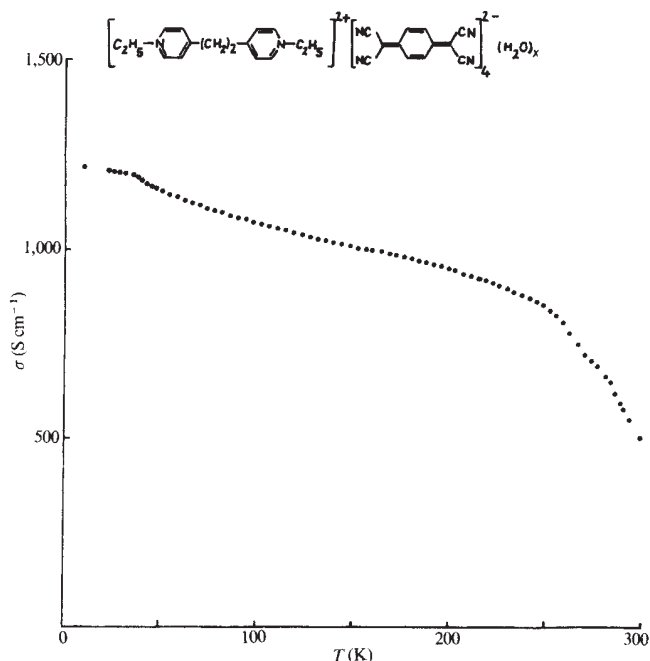


Fig. 2 Variation of the conductivity with temperature of (DEPA)(TCNQ)₄(H₂O)_x.

determined by elemental (C, H, N and O), spectral and thermogravimetric techniques (Table 1). The space group and unit cell parameters were obtained from oscillation and Weissenberg photographs using CuKα radiation (Table 2). Oscillation photographs about *b* showed strong subcell layer lines with *d*-spacings of 3.84–3.85 Å and weak layer lines with *d*-spacings of 17.4 Å in (DEPE)(TCNQ)₄(H₂O)_x, 17.1 Å in (DMPP)(TCNQ)₄(H₂O)_x and 7.68 Å in (DMPA)(TCNQ)₄(H₂O)_x. When the crystals were annealed in vacuum at ~370 K the weak layer lines could no longer be observed. In these conditions dehydration occurs with a contraction of the unit cell. The unit subcell parameters of (DMPP)(TCNQ)₄(H₂O)_x, refined on a Hilger and Watts four-circle diffractometer, were *a* = 12.93(2), *b* = 3.85(1), *c* = 27.50(2) Å, β = 108.1(2)°, *U* = 1,301 Å³ before annealing and *a* = 12.84(1), *b* = 3.84(1), *c* = 27.31(6) Å, β = 106.6(3)°, *U* = 1,290 Å³ after. In these quasi-one-dimensional conductors the weak layer lines are attributed to the H₂O lattice, commensurate in (DMPA)(TCNQ)₄(H₂O)_x and incommensurate in (DEPE)(TCNQ)₄(H₂O)_x and (DMPP)(TCNQ)₄(H₂O)_x. Variations in the intensities of the weak layer lines may indicate incomplete hydration and therefore a random occupancy of the available sites within a well defined H₂O lattice.

In (DEPE)(TCNQ)₄(H₂O)_x, (DMPP)(TCNQ)₄(H₂O)_x and (DMPA)(TCNQ)₄(H₂O)_x the cation lattice is inherently disordered which is characteristic of all highly conducting TCNQ salts with stoichiometries other than 1:1. The TCNQs stack plane-to-plane within homologous columns^{6,7} parallel to *b* with (1) a uniform mean interplanar spacing of 3.2 Å and (2) an exocyclic double bond to quinonoid ring overlap of adjacent molecules, with the longitudinal molecular staggering in a consistent direction throughout the stack. At ambient temperature the disorder inhibits a periodic coupling of the anion and cation chains and so suppresses the onset of a three-dimensional transition to a less conducting phase. At ~40 K the crystals undergo conductor-to-insulator transitions (Figs 1, 3) which are attributed to distortions of the quasi-one-dimensional TCNQ lattice. In these salts it is improbable that interactions between the conduction band electrons and the H₂O lattice, incommensurate in (DEPE)(TCNQ)₄(H₂O)_x and (DMPP)(TCNQ)₄(H₂O)_x, are sufficient to cause the Peierls distortions at

Table 1 Elemental, spectral and thermogravimetric data

Complex		C (%)	H (%)	N (%)	O (%)	$\epsilon_{395}/\epsilon_{842}$	Weight loss
(DEPE)(TCNQ) ₄ (H ₂ O) _x	Calculated*	71.65	3.52	23.51	1.31	2.2	1.5%
	Found	71.90	3.50	24.03	1.20	2.2	1.3%
(DMPP)(TCNQ) ₄ (H ₂ O) _x	Calculated†	71.31	3.57	23.77	1.36	2.2	
	Found	71.64	3.59	23.93	1.15	2.1	
(DMPA)(TCNQ) ₄ (H ₂ O) _x	Calculated‡	70.99	3.44	24.05	1.53	2.2	
	Found	71.01	3.23	24.80	1.36	2.1	

Values calculated by assuming: * $x = 0.88$; † $x = 0.90$ and ‡ $x = 1.0$. The non-stoichiometry is suggested by X-ray analysis; weak layer lines with d -spacings of $b/0.22$ in (DEPE)(TCNQ) $_4$ (H $_2$ O) $_x$ and $b/0.225$ in (DMPP)(TCNQ) $_4$ (H $_2$ O) $_x$ are attributed to an incommensurate H $_2$ O lattice.

(DEPE)(TCNQ)₄(H₂O)_x = [1,2-bis(1-ethyl-4-pyridinium)ethylene]²⁺(7,7,8,8-tetracyanoquinodimethanide)₄²⁻(H₂O)_x. (DMPP)(TCNQ)₄(H₂O)_x = [1,3-bis(1-methyl-4-pyridinium)propane]²⁺(7,7,8,8-tetracyanoquinodimethanide)₄²⁻(H₂O)_x. (DMPA)(TCNQ)₄(H₂O)_x = [1,2-bis(1-methyl-4-pyridinium)ethane]²⁺(7,7,8,8-tetracyanoquinodimethanide)₄²⁻(H₂O)_x.

~40 K. If they were we should expect the transition in $(\text{DMPA})(\text{TCNQ})_4(\text{H}_2\text{O})_x$ to occur at a higher temperature, as the H_2O lattice is commensurate in this salt. Instead I suggest that the distortions arise, in part, from an ordering of the cation lattice on cooling.

The conductor-to-insulator transitions at 40 K have been suppressed, in a few instances only, by annealing the hydrated crystals in vacuum (10^{-4} torr). The conductivities increased on dehydration. For (DEPE)(TCNQ) $_4$ (H $_2$ O) $_x$, shown in Fig. 1, the conductivity decayed on thermal recycling but, nevertheless, remained metallic. For (DEPA)(TCNQ) $_4$ (H $_2$ O) $_x$, shown in Fig. 2, the conductivity was reproducible for two cooling cycles (300–10 K) and increased monotonically on cooling to below 10 K. In this low temperature region the properties depend on the degree of disorder which, in the limit, can obscure the Peierls transition by, for example, introducing a substantial density of states into the band gap as exemplified by recent work on (dimethyldiphenylphosphonium) $^+$ (TCNQ) $_2^-$ (ref. 8). The experimental data reflect an increase in the disorder on dehydration and a decrease on thermal recycling.

The electrical properties of $(\text{DEPE})(\text{TCNQ})_4(\text{H}_2\text{O})_x$, $(\text{DMPP})(\text{TCNQ})_4(\text{H}_2\text{O})_x$, $(\text{DMPA})(\text{TCNQ})_4(\text{H}_2\text{O})_x$ and $(\text{DEPA})(\text{TCNQ})_4(\text{H}_2\text{O})_x$ depend on the outgassing procedure.

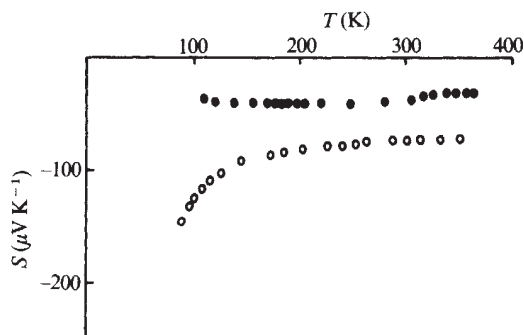


Fig. 4 Variation of the thermoelectric power with temperature of two compacted pellets of (DEPE)(TCNQ)₄(H₂O)_x, one hydrated (○) and the other dehydrated (●).

Crystals annealed in vacuum at 300–400 K have variable properties which may be attributed to the onset of decomposition, increased disorder and to loss of solvent from the lattice. Five crystals have shown metallic behaviour with $\sigma_{300\text{K}} = 100\text{--}500 \text{ S cm}^{-1}$ along *b*. Most were semiconducting with $\sigma_{300\text{K}} = 50 \text{ S cm}^{-1}$. The metallic crystals are believed to be more disordered and less hydrated than the semiconducting crystals. For a correlated semiconductor the thermoelectric power, *S*, is given⁹ by

$$S = -k/e \ln[2(1-\rho)/\rho] \mu V K^{-1}$$

where ρ is the ratio of carriers to available TCNQ sites. Assuming that charge transfer is commensurate with the chemical formulae presented for the isostructural series, $\rho = 0.5$ and $S = -59.8 \mu\text{V K}^{-1}$. For (DEPE)(TCNQ) $_4$ (H $_2$ O) $_x$ this theoretical value is consistent with the experimental data at saturation for the hydrated sample only (Fig. 4). The semiconducting behaviour probably arises from a pinning of the charge density waves by the H $_2$ O 'impurities'. When the pinning is destroyed, by dehydration, the conductivity is enhanced and may become metallic. Thus in the dehydrated sample the thermoelectric

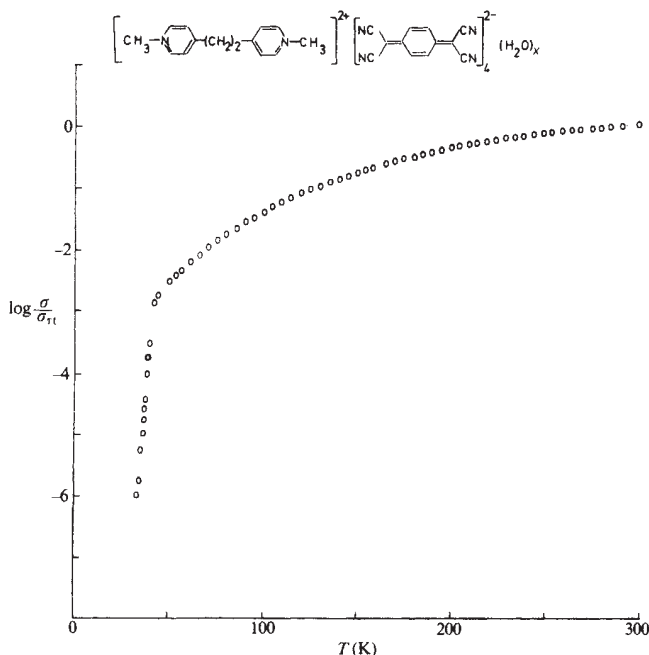


Fig. 3 Variation of the normalized conductivity with temperature of $(\text{DMPA})(\text{TCNQ})_4(\text{H}_2\text{O})_x$.

Table 2 Unit subcell parameters of the strong layer lines and *d*-spacings of the weak layer lines of the isostructural series of conducting TCNO salts

Complex	Space group	<i>a</i> (Å)	<i>b</i> (Å)	<i>c</i> (Å)	β (°)	<i>d</i> -spacing (Å)
(DEPE)(TCNQ) ₄ (H ₂ O) _x	P2 ₁ /c	12.72	3.84	27.67	107.8	17.4
(DMPP)(TCNQ) ₄ (H ₂ O) _x	P2 ₁ /c	12.93	3.85	27.50	108.1	17.1
(DMPA)(TCNQ) ₄ (H ₂ O) _x	P2 ₁ /c	12.90	3.84	27.34	107.4	7.68

The weak layer lines are attributed to the H₂O lattice.

e.m.f. is significantly lower and the properties are no longer consistent with those of a correlated semiconductor. On dehydration there is no significant change in the TCNQ stacking characteristics.

All four salts have crystallographically ordered anhydrous modifications. In these the TCNQs are stacked plane-to-plane in stoichiometric groups of four, in a direction parallel to the length of the cation, with no direct overlap between adjacent tetrads. This periodic distortion arises from a periodic coupling of the anion and cation chains. The conductivities are significantly lower with $\sigma_{300K} \sim 0.002 \text{ S cm}^{-1}$ along b , the direction of stacking. Conversely in the crystallographically disordered modifications the periodic interchain coupling is destroyed by cation disorder and homologous columns of TCNQs occur. In both types of modification the electrical properties are consistent with those of related complexes in which the TCNQ stacking characteristics are similar.

I thank Dr S. C. Wallwork for the refined subcell parameters of $(\text{DMPP})(\text{TCNQ})_4(\text{H}_2\text{O})_x$ and Dr G. Owen for helpful discussions. I also thank G. H. Cross for assistance with the conductivity measurements and the Chemical Society Research Fund and the SRC for support.

Received 14 November 1980; accepted 9 February 1981.

- Jerome, D., Mazaud, A., Ribault, M. & Bechgaard, K. *J. phys. Lett., Paris* **41**, L95 (1980).
- Bechgaard, K., Jacobsen, C. S., Mortensen, K., Pedersen, H. J. & Thorup, N. *Solid St. Commun.* **33**, 1119 (1980).
- Ashwell, G. J., Eley, D. D. & Willis, M. R. *Nature* **259**, 201 (1976).
- Ashwell, G. J. *Phys. St. Solidi* **86b**, 705 (1978).
- Somoano, R. B. *et al. Phys. St. Solidi* **81b**, 281 (1977).
- Ashwell, G. J. *Proc. int. Conf. on Low Dimensional Synthetic Metals*, Helsinki (1980).
- Wallwork, S. C. *Proc. Conf. on Organic and Biological Semiconductors*, Nottingham (1980).
- Ashwell, G. J., Allen, D. W., Cross, G. H., Kennedy, D. A. & Nowell, I. W. (in preparation).
- Chaikin, P. M. & Beni, G. *Phys. Rev.* **B13**, 647 (1976).

High abundances of long-chain normal alkanolic acids in Antarctic soil

Genki Matsumoto*, Tetsuya Torii† & Takahisa Hanya*

* Department of Chemistry, Faculty of Science, Tokyo Metropolitan University, Fukazawa 2-1-1, Setagaya-ku, Tokyo 158, Japan

† Chiba Institute of Technology, Tsudanuma, Narashino, Chiba Prefecture 275, Japan

Studies of geochemical samples have shown that fatty acids are widely distributed in sediments ranging in geological age from Recent to Precambrian¹⁻⁹. Most *n*-alkanoic acids found in soils and Recent sediments show similarities in their distributions, such as a marked predominance of even-carbon numbered acids and a bimodal distribution with two maxima in carbon chain length of C_{14} to C_{18} and C_{24} to C_{28} (refs 1, 2, 4-7). Long-chain *n*-alkanoic acids ($\geq \text{C}_{20}$) predominate in the waxes of vascular plants^{10,11}, whereas short-chain *n*-alkanoic acids ($< \text{C}_{20}$) occur in the lipids of all living organisms. Long-chain *n*-alkanoic acids found in soils and Recent sediments are thus believed to arise from vascular plants¹²⁻¹⁴. In Antarctica, vascular plants are only found in the Antarctic Peninsula¹⁵. Organic constituents on this continent are thus expected to be considerably different from those in the temperate and tropical zones. We report here the presence of very long-chain *n*-alkanoic acids extending to C_{40} , having a small even-carbon predominance, and comprising a large proportion of the fatty acids in Antarctic soils. They apparently originate in the Beacon Sandstone.

During the 1967-77 austral summer, soil samples were collected from the dry valleys of Victoria Land in Antarctica (Fig. 1) with a small stainless-steel scoop, wrapped in Teflon sheet and stored at temperatures below 0°C until analysis. The

procedures for fatty acid analysis have been described in detail elsewhere¹⁶. Briefly, wet samples ($\sim 50 \text{ g}$) were extracted with ethyl acetate after saponification with 0.5 M potassium hydroxide/methanol solution. Each extract was chromatographed on a silica gel column ($160 \times 4\text{-mm}$ i.d., 100 mesh, 5% water). After elution with two column volumes of hexane, the eluate (fatty acid fraction) obtained with three column volumes of benzene/ethyl acetate (95/5) was methylated with 15% boron trifluoride/methanol solution (3 ml, 80°C , 2 h). The structures of the methyl esters of fatty acids were determined using a combined Shimadzu LKB 9000 gas chromatograph-mass spectrometer in conditions given in Fig. 2.

Fatty acids were identified by comparing their retention times and mass spectra with those of authentic compounds and homologues. Quantitative acid data were obtained by measuring the peak area on the mass fragmentogram or gas chromatogram (total ion current). A blank test to check contamination throughout the experiments indicated possible contaminants (Fig. 2), but they were present in such low amounts that they did not affect the analytical results for long-chain *n*-alkanoic acids. Based on four replicate addition experiments using authentic compounds, the percentage recoveries of *n*-alkanoic acids, C_{14} , C_{16} , C_{18} , C_{20} and C_{22} , were 88 (s.d. 4.0), 91 (8.0), 87 (3.5), 88 (3.7) and 107 (13), respectively.

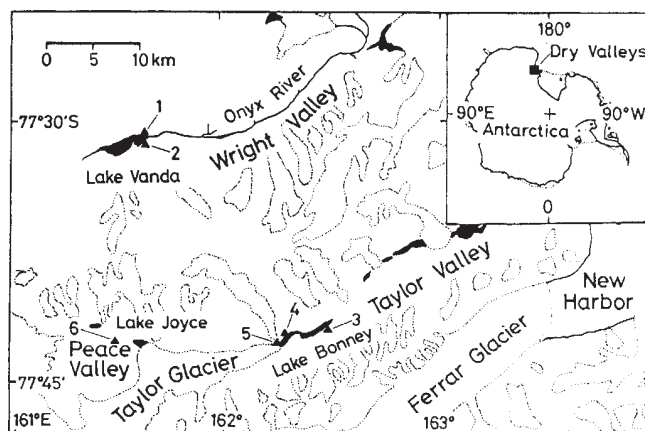


Fig. 1 Sampling locations in the dry valleys of Victoria Land in Antarctica. 1, Vanda-1; 2, Vanda-2; 3, Bonney, east lobe; 4, Bonney, west lobe-1; 5, Bonney, west lobe-2; 6, Joyce.

The analytical results from six soil samples are summarized in Table 1. The total acid content of the soils is fairly low, ranging from 0.16 to $3.7 \mu\text{g}$ per g dry soil, presumably reflecting the low organic carbon contents (0.14 – 0.71 mg C per g)¹⁶. *n*-Alkanoic acids ranging from C_8 to C_{40} were found with a bimodal distribution maximizing at C_{16} and C_{28} , together with branched (iso and ante-iso) and unsaturated acids (Fig. 2). As far as we know, this is the first report of the occurrence of such very long-chain *n*-alkanoic acids ($> \text{C}_{38}$) in soils and modern and ancient sediments.

Surprisingly, long-chain *n*-alkanoic acids comprised most of the fatty acids, ranging from 30.2 to 81.9% with an average of $53.8 \pm 18.7\%$ at 90% confidence limits, with a small even-carbon predominance. Thus the ratios of long- to short-chain *n*-alkanoic acids are very high (2.7 ± 1.7) and comparable with the value estimated for a forest soil (2.3)¹¹. The carbon preference indices, CPI_{AL} , CPI_{AH} and CPI_{AT} , calculated over the ranges C_{13} – C_{19} , C_{19} – C_{33} and C_{13} – C_{33} according to Kvenvolden^{1,4}, ranged from 4.5 to 13 , 1.4 to 3.6 and 1.7 to 5.7 , respectively. In general, these values are lower than those in common biological materials. In particular, the CPI_{AH} and CPI_{AT} values of the three Bonney samples are extremely low, and are thought to be the smallest values ever reported for soils, although CPI_{A} values close to unity have been found in some

Table 1 Fatty acids found in soils collected from the dry valleys of Victoria Land in Antarctica

Locality	Nature of soil	Range	Maximum at carbon number*	Concentration (µg per g)	Composition (%)				Ratio long/short	CPI _{AL}	CPI _{AH}	CPI _{AT}
					Short	Long	Branched	Un-saturated				
Wright Valley												
Vanda-1	OY, silt	C ₁₀ -C ₃₂	<u>C₁₆</u> , C ₂₄	3.7	31.1	31.6	5.4	31.9	1.0	13	3.6	5.7
Vanda-2	DYB, sand	C ₈ -C ₃₄	<u>C₁₆</u> , C ₂₈	2.4	30.4	46.7	9.4	13.5	1.5	7.8	3.2	4.3
Taylor Valley												
Bonney, east lobe	GO, sand	C ₈ -C ₄₀	C ₁₆ , <u>C₂₈</u>	0.94	17.1	52.2	5.8	24.9	3.1	6.1	1.7	2.2
Bonney, west lobe-1	OY, sand	C ₈ -C ₄₀	C ₁₆ , <u>C₂₈</u>	0.43	16.7	80.2	1.3	1.8	4.8	4.8	1.4	1.7
Bonney, west lobe-2	OY, sand	C ₈ -C ₃₇	C ₁₆ , <u>C₂₈</u>	0.91	15.5	81.9	1.4	1.2	5.3	4.5	1.5	1.8
Pearce Valley												
Joyce	YB, sand	C ₈ -C ₃₅	<u>C₁₆</u> , C ₂₈	0.16	56.1	30.2	9.0	4.7	0.54	5.3	1.5	3.1
Average, 90% confidence limits				1.4 ± 1.1	27.8 ± 12.8	53.8 ± 18.7	5.4 ± 2.9	13.0 ± 10.6	2.7 ± 1.7	6.9 ± 2.6	2.2 ± 0.81	3.1 ± 1.3

Carbon preference indices, CPI_{AL}, CPI_{AH} and CPI_{AT}, values were calculated from the carbon chain lengths of C₁₃-C₁₉, C₁₉-C₃₃ and C₁₃-C₃₃, respectively. Short, short-chain *n*-alkanoic acids (C₁₃-C₁₉). Long, long-chain *n*-alkanoic acids ($\geq$ C₂₀). Branched, branched acids (iso and ante-iso, C₁₃-C₁₇). Unsaturated, unsaturated acids (C₁₆ and C₁₈). OY, olive yellow; DYB, dull yellowish brown; sand, coarse and fine sands; GO, greyish olive. YB, yellowish brown.

* The major distribution maximum is underlined.

sedimentary rocks and shales^{1,2,4}. However, we found no long-chain *n*-alkanoic acids with small CPI_A values, which make up a high proportion of the fatty acids in lake waters, sediments and epibenthic algae, that is, blue-green algae (mostly *Oscillatoria* spp., *Phormidium* spp. and *Calothrix* spp.), in samples collected from Lakes Vanda, Bonney and Joyce^{16,17}.

Little is known about the occurrence of long-chain *n*-alkanoic acids other than in the waxes of vascular plants^{10,11} and bees waxes^{18,19}. There are no vascular plants and bees in the areas studied and these long-chain *n*-alkanoic acids are not due to contamination during analysis. There seem to be at least three possible sources of these unusual *n*-alkanoic acids.

(1) An input of aeolian dust, because air-borne pollutants such as dichlorodiphenyl-trichloroethane are found in Antarctic snow²⁰. Some aeolian dusts collected at sea near land areas contain long-chain *n*-alkanoic acids in large relative abundances, but their CPI_A values (6.1-22.0) are considerably higher than ours²¹. Furthermore, the fatty acid ($\geq$ C₁₀)/*n*-alkane ($\geq$ C₁₅) ratios of the aeolian dusts (0.088-0.50)²¹ are much lower than those of the Antarctic soils (31-110, ref. 16; Matsumoto *et al.*, unpublished results). Thus aeolian dust is not the main source of our fatty acids.

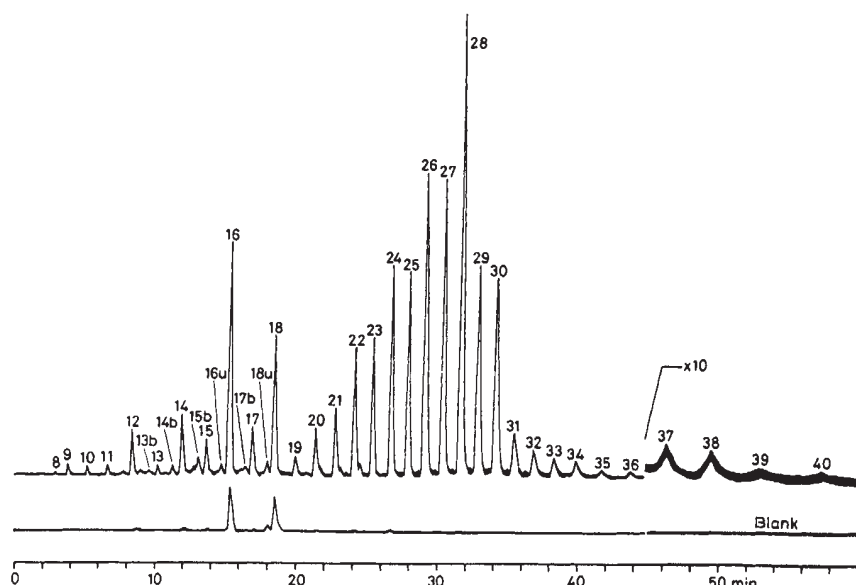
(2) Microbial oxidation of *n*-alkanes, because many kinds of microorganism including bacteria and fungi are found in soils in the dry valley areas²². Although *n*-alkanes ranging from C₁₅ to C₃₇, with the most dominant peak at C₂₃ or C₂₇, were found at very low total amounts, between 0.0045 and 0.055 $\mu\text{g per g}$ dry

soil (ref. 16; Matsumoto *et al.*, unpublished results), they could not give our fatty acid profile after microbial oxidation (Fig. 2, Table 2). Thus *n*-alkanes are also an unlikely source.

(3) Glacial erosion of sedimentary rocks containing organic materials may have contributed long-chain *n*-alkanoic acids or precursors, as long-chain *n*-alkanoic acids having small CPI_A values, and high long/short chain ratios have been found in sedimentary rocks and shales of Palaeozoic to Cenozoic age from the US^{1,2,4}. In our case, the dry valley areas are mainly composed of sedimentary, metamorphic and igneous rocks of Precambrian to Cretaceous age. It is well known that the Beacon Sandstone, composed of sandstone, siltstone, conglomerate and carbonaceous beds, and including various fossils such as coals and silicified woods, is widely distributed in the dry valley areas²³. Much of the sedimentary organic carbon in the Ross Sea is reported to be derived from the rocks being eroded by glaciers on the Antarctic continent²⁴. It is therefore likely that the long-chain *n*-alkanoic acids found in Antarctic soils come from the Beacon Sandstone. On the other hand, the short-chain acids, which are similar to those of common biological materials, can be mainly attributed to recent biological sources. Further samples of Antarctic sedimentary rocks including fossils will be collected for the analysis of organic constituents.

We thank Mr Y. Tanaka, Chiba Institute of Technology, the Antarctic Division, DSIR, New Zealand, the US NSF and the National Institute of Polar Research, Japan, for their support in collecting samples.

Fig. 2 Mass fragmentogram (m/e 74) of the fatty acid fraction obtained from soil sample collected from Victoria Land in Antarctica (Bonney, west lobe-1). Conditions: silanized glass column (200 cm $\times$ 3 mm i.d.) packed with 1% silicone OV-1 on 80-100-mesh Chromosorb W AW DMCS; temperature, injector 300 °C, column programmed from 100 to 290 °C at 6 °C min⁻¹, molecular separator 300 °C, ion source 330 °C; carrier gas, helium at 30 ml min⁻¹; acceleration voltage, 3.5 kV; electron energy, 20 eV; multiplier gain, 3. Arabic figures at the peaks indicate the carbon chain length of fatty acids. 13b, 14b, 15b and 17b are branched (iso + ante-iso) C₁₃, C₁₄, C₁₅ and C₁₇ acids, respectively. 16u and 18u are unsaturated C₁₆ and C₁₈ acids, respectively.



Received 25 March 1980; accepted 10 February 1981.

1. Kvenvolden, K. A. *Nature* **209**, 573–577 (1966).
2. Kvenvolden, K. A. *J. Am. Oil Chem. Soc.* **44**, 628–636 (1967).
3. Han, J. & Calvin, M. *Nature* **224**, 576–577 (1969).
4. Kvenvolden, K. A. in *Advances in Organic Geochemistry 1966* (eds Hobson, G. D. & Speers, G. C.) 335–366 (Pergamon, Oxford, 1970).
5. Brown, F. S., Baedecker, M. J., Nissenbaum, A. & Kaplan, I. R. *Geochim. cosmochim. Acta* **36**, 1185–1203 (1972).
6. Eglinton, G., Maxwell, J. R. & Philp, R. P. in *Advances in Organic Geochemistry 1973* (eds Tissot, B. & Biener, F.) 941–961 (Editions Technip, Paris, 1974).
7. Matsuda, H. & Koyama, T. *Geochim. cosmochim. Acta* **41**, 777–783 (1977).
8. Van Vleet, E. S. & Quinn, J. G. *Geochim. cosmochim. Acta* **43**, 289–303 (1979).
9. Volkman, J. K., Johns, R. B., Gillan, F. T. & Perry, G. J. *Geochim. cosmochim. Acta* **44**, 1133–1143 (1980).
10. Kolattukudy, P. E. *Lipids* **5**, 259–275 (1970).
11. Matsuda, H. & Koyama, T. *Geochim. cosmochim. Acta* **41**, 1825–1834 (1977).
12. Morrison, R. I. in *Organic Geochemistry. Methods and Results* (eds Eglinton, G. & Murphy, M. T. J.) 558–575 (Springer, Berlin, 1969).
13. Cranwell, P. A. *Chem. Geol.* **14**, 1–14 (1974).
14. Brooks, P. W. et al. *Chem. Geol.* **18**, 21–38 (1976).
15. Greene, S. W. et al. in *Terrestrial Life of Antarctica*, Antarctic Map Folio Series 5 (ed. Bushnell, V. C.) 1–24 (American Geographic Society, New York, 1967).
16. Matsumoto, G., Torii, T. & Hanya, T. *Mem. natl. Inst. Polar Res. Spec. Issue No. 13*, 103–120 (1979).
17. Matsumoto, G. & Hanya, T. *Antarctic Rec.* **58**, 81–88 (1977).
18. Holloway, P. J. *J. Am. Oil Chem. Soc.* **46**, 189–190 (1969).
19. Tulloch, A. P. & Hoffman, L. L. *J. Am. Oil Chem. Soc.* **49**, 696–699 (1972).
20. Peel, D. A. *Nature* **254**, 324–325 (1975).
21. Simoneit, B. R. T. & Eglinton, G. in *Advances in Organic Geochemistry 1975* (eds Campos, R. & Goñi, J.) 415–430 (Enadinsa, Madrid, 1977).
22. Cameron, R. E., King, J. & David, C. N. in *Antarctic Ecology Vol. 2* (ed. Holdgate, M. W.) 702–716 (Academic, London, 1970).
23. Warren, G. & Gunn, B. M. in *Geology of Mawson-Mulock Area, Victoria Land, Antarctica* (Dept of Scientific and Industrial Research, Wellington, 1961).
24. Sackett, W. M., Poag, C. W. & Eadie, B. J. *Science* **185**, 1045–1047 (1974).

α -Particle radioactivity of hot particles from the Esk estuary

E. I. Hamilton

Institute for Marine Environmental Research, Prospect Place,
The Hoe, Plymouth PL1 3DH, UK

Transuranium radionuclides (Pu, Am and Cm) present in effluents discharged into the north-east Irish Sea by British Nuclear Fuels Limited, Windscale, Cumbria, UK, are found in sediment and biota of the Esk estuary ~10 km to the south. The site of the present investigation was at Newbiggin and the materials examined were suspended particulate debris samples at the sea surface, bottom sediments and some forms of biota collected in September 1977. It is shown here that hot particles (defined as small volumes of material emitting α particles recorded in a dielectric detector as dense clusters of tracks from a common origin) found in the estuary are likely to be original effluent debris derived from the processing of Magnox uranium fuel elements and not formed *in situ* as a result of natural processes common to the estuary.

Methods of analysis described elsewhere^{1–4} use a new dielectric α -track detector, CR-39 (ref. 5), in association with conventional chemical separation techniques and α -particle spectrometry using surface barrier detectors. The CR-39 method involves placing a sample in contact with a sheet of detector and after a suitable period of exposure the α particles recorded are developed by etching at 80 °C in 6.0 M NaOH for 3 h. After etching, the detector is viewed with the naked eye and α tracks associated with hot particles can be easily observed against a background of randomly distributed tracks; analysis of α -particle track distribution is then determined using a conventional light microscope with $\times 5$ or $\times 16$ objectives and $\times 10$ oculars; quantitative analysis of track densities is determined using a Quantimet image analyser.

In the Esk estuary (at the Newbiggin rail viaduct), after slack water and when the flood tide is just starting, the water becomes very turbid³; for a short period of time (15–45 min depending on the energy of the tide), fine-grained (<100 μ m) bottom deposited silts become resuspended into the water column (0.1–0.5 m deep) and just reach the surface, although in calm water conditions the top ~1 mm of water is often clear. In calm conditions the paper side of a sheet (30 $\times$ 30 cm) of Whatman

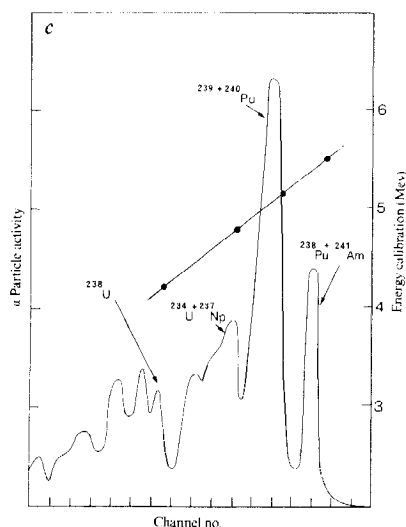
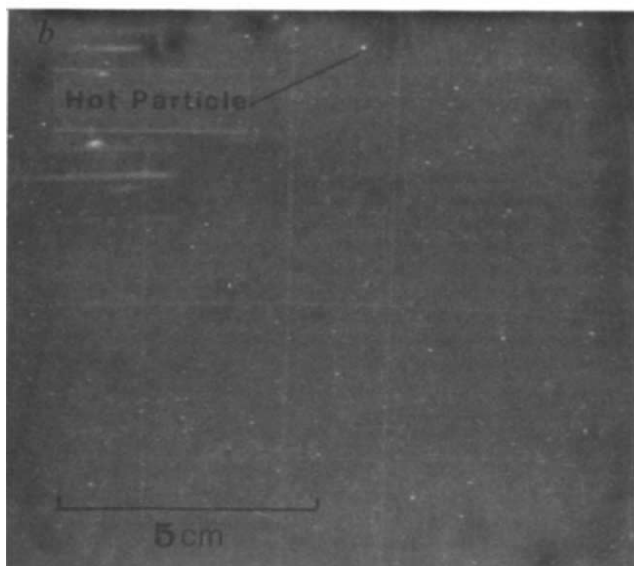
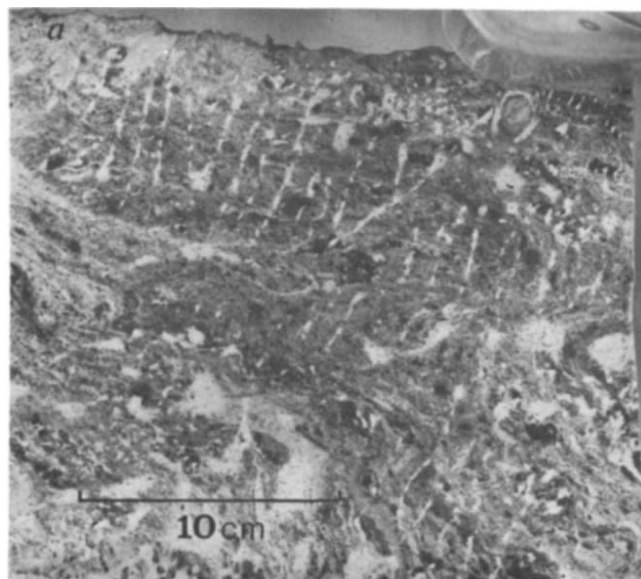


Fig. 1 Identification of the distribution of hot particles in sediment using CR-39 and surface barrier α spectrometry. *a*, Seawater surface sediment absorbed onto Whatman Benchkote. *b*, Etched CR-39 illustrating (white spots) sites of hot particles in *a*. *c*, Thick source α spectra of a single hot particle showing the presence of U, Pu and Am radionuclides.

Table 1 $^{239+240}\text{Pu}$: ^{238}Pu activity ratios, percent abundance of $^{238+239+240}\text{Pu}$ and ^{241}Am in various materials from the Esk estuary, Cumbria, UK and adjacent areas collected in 1977

Material	α particle activity ratio $^{239+240}\text{Pu}/^{238}\text{Pu}$	α emitting radionuclides (% abundance)		Mean Pu/Am (α activities)	Ref.
		$^{238+239+240}\text{Pu}$	^{241}Am		
Newbiggin silts ($n=8$)	3.6 ± 0.4	46.3	53.7	0.9 ± 0.2	3
Top 2 cm sediment core 1176 Newbiggin	3.4 ± 0.4	57.4	42.6	1.4 ± 0.4	3
Regional sediments ($n=40$)	4.3 ± 0.5	53.0	47.0	1.1 ± 0.3	21
Benthic animals ($n=56$)	4.2 ± 0.3	46.2	53.8	0.9 ± 0.5	3
Fish ($n=5$)	3.9 ± 0.5	54.7	45.3	1.2 ± 0.2	21
Mean	3.9 ± 0.3	—	—	1.1 ± 0.2	
<i>Fucus vesiculosus</i>					
Ravenglass	3.7	86.7	13.3	6.5	3
Newbiggin	3.8	65.8	34.2	1.9	21

Data obtained following chemical separation and surface barrier α spectrometry.

Benchkote is touched onto the sea surface, just penetrating the clear surface layer of water if present, to which particulates recently resuspended into the water column adhere and are removed. Visible swirls of particulate debris present on the sea surface are retained by this technique as shown in Fig. 1a. Recently exposed bottom sediments are sampled similarly and if no pressure is applied to the paper, only the top layer of sediment (~ 1 mm) is removed. The Benchkote is then allowed to dry until just damp when it is placed against a sheet of CR-39 and exposed for between 2 h and 10 days. After the detector is etched two features are visible to the naked eye (Fig. 1b); first, the presence of diffuse opalescent markings which reproduce the macrostructure of the adhered sediment, and second, randomly distributed white spots ($<100 \mu\text{m}$ mean particle diameter) which represent the intense activity originating from hot spots. The position of the hot particles in the detector is noted and the corresponding site on the Benchkote identified. A small area of Benchkote enclosing a hot particle is then cut out; typically the level of diffuse activity associated with the area of paper removed is insignificant and all the activity originates from the hot particle. Briefly the following analytical techniques are then used. (1) The total number of α tracks from each separated particle is determined using the Quantimet analyser calibrated for α -particle detection efficiency against a standardized zinc sulphide screen counter⁴. When the track density is too intense with the result that individual tracks cannot be recognized, the total activity of the particle is determined by the zinc sulphide screen method; typically particle activities are between 0.1 and 3 pCi for total α emissions. (2) The hot particle attached to the Benchkote is then mounted in an energy calibrated surface barrier α spectrometer and the α spectrum determined. In the samples investigated so far ($n=28$), four of the spectra were partially degraded; these particles were dissolved in hot hydrochloric acid together with ^{236}Pu and ^{243}Am yield monitors, the detrital Benchkote removed and the leachate, after adjustment to pH of a plating solution, electroplated onto a polished stainless steel planchet. The total α activity recorded after electroplating agreed with that obtained by the zinc sulphide method and validates hydrochloric acid being used instead of nitric acid³ as a leaching agent to reduce the degradation of the Benchkote filter paper. None of the particles examined gave a totally degraded α spectrum and using the procedures described by Evans⁶ individual energies were identified. α -particle composition of the particles varied considerably; Pu and Am were always present, curium radionuclides were not always present; in six samples low energy emissions at ~ 4.8 meV and 4.2 meV were

observed and are tentatively identified as originating from $^{234}\text{U} + ^{237}\text{Np}$ and ^{238}U respectively.

When electroplating techniques are used without prior chemical separation of Pu and Am, because of energy overlap the radionuclides determined are $^{239+240}\text{Pu}$, $^{238}\text{Pu} + ^{241}\text{Am}$, $^{243+244}\text{Cm}$ and ^{242}Cm ; compared with standard chemical techniques³ no separation of ^{238}Pu from ^{241}Am is obtained. To convert the data into conventional terms the ^{238}Pu component had to be identified. Using chemical separation methods, together with published data, a $^{238}\text{Pu}/^{239}\text{Pu}$ ratio of 3.9 ± 0.3 determined on a large number of sediments and biota collected in September 1977 from the area, was used to determine the ^{238}Pu component of the $^{238}\text{Pu} + ^{241}\text{Am}$ α -spectra peaks. The activity of individual α emitters from Pu and Am was then calculated in terms of the percentage abundance for total Pu (α) and Am; the data are given in Table 1 and Fig. 2 and show that although the mean Pu/Am ratio for sediments and biota of the area is 1.1 ± 0.2 , the sea surface particulates showed a much greater range (Pu/Am, 3.2 ± 3.0) with a distinct trend towards an enrichment in Pu; none of the samples was significantly enriched in Am. A preliminary examination of hot particles ($n=10$) from the top 10 cm of deposited sediments gave a Pu/Am ratio of 0.9 ± 0.2 ; however, the morphology of these particles differs from those collected near the sea surface and will be discussed elsewhere. Note, however, that there are fewer hot particles present in buried sediments, and the specific activity is less. Data for *Fucus vesiculosus* indicates some degree of Pu enrichment which may be related to the entrainment of particles on fronds³.

In the absence of any quantitative data which describes the isotopic composition of transuranics, together with their chemical and physical form in fresh effluent released by BNFL, they cannot be identified in the Esk. However, circumstantial evidence allows some interpretation of the data. Using data presented by Nair⁷ it is calculated that the specific α radioactivity of aged Magnox fuel is $\sim 0.2 \text{ Ci g}^{-1}$. Measurement of particle size, using a technique based on the diameter of hotspots⁴ shows that the highest activity is usually associated with particles with a mean particle diameter of $<20 \mu\text{m}$; assuming that the density of these particles is ~ 10 they would have a maximum total α activity of $1\text{--}10 \mu\text{Ci g}^{-1}$, and therefore they do not consist of unprocessed uranium fuel. In the Esk estuary undisturbed sediments contain significant levels of short-lived radionuclides such as $^{95}\text{Zr-Nb}$ and ^{242}Cm , hence seaborne transit times for effluent from the outfall to the Esk are rapid enough for short-lived radionuclides to reach the area; indeed the estimated transit time is $<1\text{--}3$ months⁸, hence the presence or absence of water transported particulate debris at any single sampling time will depend on the previous weather conditions

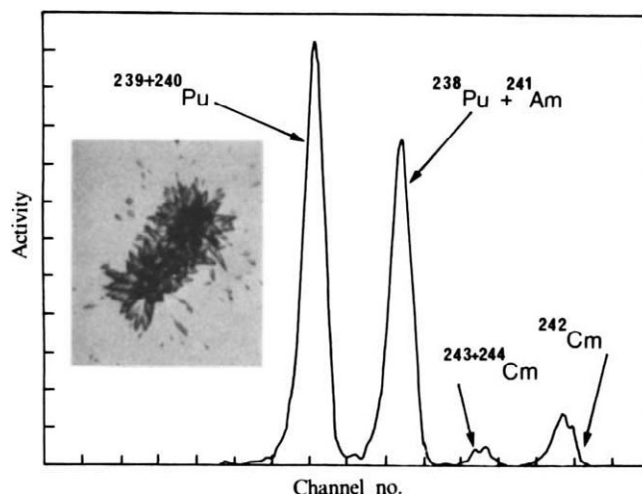


Fig. 2 Photograph and direct α spectra for a hot particle containing Pu, Am and Cm radionuclides sampled at the sea surface.

Table 2 Mean monthly discharges (Ci yr^{-1}) for Pu and Am α radioactivity) from the uranium reprocessing plant together with Pu/Am ratios⁹

Year	$^{238}\text{Pu} + ^{239+240}\text{Pu}$	^{241}Am	Pu/Am
1977	984	96	10.3
1976	1272	324	3.9
1975	1200	984	1.2
1974	1248	3192	0.4

common to the region. Table 2 provides data for the Pu/Am ratio, derived from mean monthly discharge data which, in 1977⁹ was 10.3; in previous years it was significantly lower. The 1977 Pu/Am ratio for Esk sediments and biota (see Table 1 and Fig. 3) can be matched with 1975 releases, thus requiring a delay of ~2 yr before they become characteristic for the Esk. No data are available for the monthly discharges of Pu or Am released in the effluent, but from data presented by Baxter¹⁰ it seems probable that maximum releases occur during September–December of each year. Other data³ show that it was not until 1979 that sediment and biota of the Esk estuary showed enhanced levels of Pu, supporting the concept of a delay period ~2 yr before 1977 effluents became dominant in the region. Recently¹¹ it has been suggested, on the basis of derived sedimentation rates for Esk sediments, that there is a finite time lag of 2–3 yr between the plutonium discharges from BNFL and the arrival of plutonium in these sediments. Fundamental differences exist in the determination of the apparent time lag and are discussed elsewhere¹². Hence on this basis surface sediment of the Esk in 1977 would not be expected to reflect isotope ratios for current releases; through rapid transport of suspended particulates in the water column such debris could be present but its contribution to bulk deposited surface sediments would be small and difficult to identify. Note that, compared with sediments, uptake of Pu and Am by biota would be expected to reflect isotope ratios in seawater common to an area when the samples were taken.¹³

Six of the spectra obtained from the hot particles showed the presence of low energy emissions at 4.2 MeV and 4.8 MeV; for other samples the energy cutoff below the $^{239+240}\text{Pu}$ peak was sharp. These energies could arise from $^{234}\text{U} + ^{237}\text{Np}$ and ^{238}U . In 1978, 16 Ci of ^{237}Np was discharged by BNFL¹⁴ (no data are available for 1977) which should be detectable as it is similar to the discharge level for ^{242}Cm (15 Ci); it is assumed that ^{237}Np has always been a component of the BNFL effluent. Recent measurements confirm the presence of ^{237}Np in Esk silts at levels comparable with those found for Cm radionuclides. ^{237}Np in *Porphyra* from Sellafield, Cumbria is reported¹⁶ at a similar level to the ^{242}Cm abundance. Use of the delayed neutron method for uranium assay has shown² that in spite of a discharge of considerable quantities of uranium in BNFL effluent (for example, 10,936 kg in 1977)¹⁴ the levels for uranium in the Esk sediments are normal. Recent measurements¹⁵ have shown that the isotopic composition of uranium in Esk sediments does not show the presence of significant levels of depleted uranium despite problems associated with the production of ^{234}U from the decay of ^{238}Pu . However, the hot particles from surface water of the Esk seem to contain α emitters which would be compatible with the presence of U and ^{237}Np . The spectra do not contain ^{226}Ra or daughter products or thorium isotopes which could be attributed to natural sources. Therefore, the presence of the (U–Np)–Pu–Am–Cm radionuclides in some hot particles is interpreted as evidence that they represent original effluent materials. The lifetime for uranium-bearing particles in the sea is not known. A 4×0.5 mm uranium rod placed in seawater (oxygenated or reduced) decomposes within a few weeks to give dissolved species and finely disseminated uranium oxides which then slowly dissolve. It seems unlikely that any uranium-rich particle would survive intact for very long in the sea, thus they are unlikely to be present in deposited sediments, but could occur in the water column as a result of recent transport.

Uranium behaves in a conservative manner in seawater and using levels for ^{137}Cs (also conservative) in seawater off Wind-scale as a measure of effluent levels, it is calculated that if the 1977 discharges of U were totally dissolved in the sea this would give rise to a concentration of $0.03 \mu\text{g U l}^{-1}$ for offshore water, and $0.1 \mu\text{g U l}^{-1}$ for near-shore water, which would be difficult to detect in the presence of normal levels of $\sim 3 \mu\text{g U l}^{-1}$ for seawater².

Spent Magnox fuel rods are usually kept submerged in tanks held at pH 11.5, containing only trace quantities of Cl^- , CO_3^{2-} and SO_4^{2-} ions to prevent solution of the Magnox; usually imperfection and impurities cause some uranium fuel loss and this accumulates on the bottom of the tanks. Before recent improvements in waste treatment, the final low activity tanks were maintained at pH 2–3 which would precipitate iron, present as a contaminant, and coprecipitate the actinides¹⁷. Before discharge into the sea the effluent is treated with ammonium hydroxide which will predispose any remaining actinides in solution to assume less soluble forms (that is, colloids). Magnesium derived from the Magnox cladding is also likely to be present in the effluent. The released effluent has a complex composition and apparently contains particulate phases, some of which would dissolve in the sea (when dispersed) according to individual compositions. Radionuclide compositions of particles found in the Esk are similar to those of the original effluent before discharge, therefore are highly unlikely to have been formed *in situ* in the Esk. In the sea, magnesium hydroxide particles are likely to dissolve slowly while labelled iron particles would be transported further, and then become subject to solution processes in the organic-rich surface sediments of the Esk as a result of geochemical recycling. In the Esk interactions with freshwater and flocculation processes which occur at the salt–freshwater interface would assist dissolution of particles.

Dissolution of hot particles is supported by the form of the particulates. Hot particles with well defined form such as shown in Fig. 2 have been found in surface sediment and water, but are not always associated with the low-energy α emitters. In buried sediments occasional hot particles are present, but usually of lower activity. They often tend to have dimensions of ~ 10 (a and b axis) $\times 30$ (c axis) μm and would not be inconsistent with

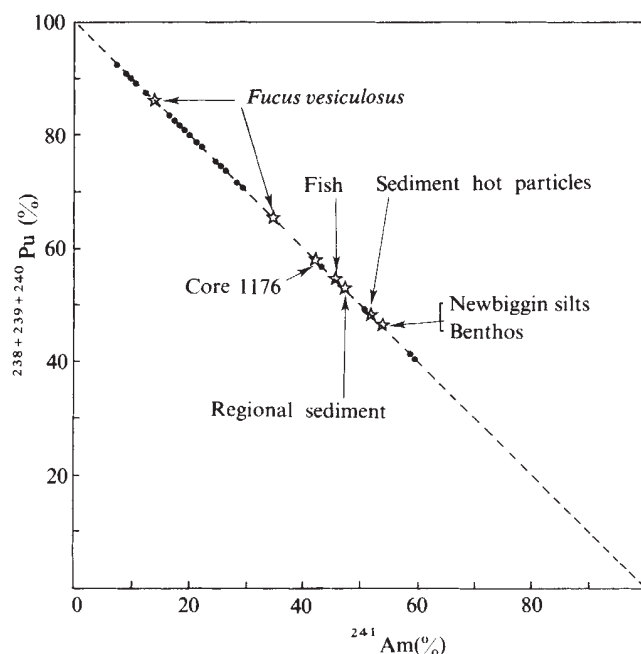


Fig. 3 Ratio Pu/Am for sea surface particulates and other material (see Table 1).

the gradual solution of the surface water particles shown in Fig. 2. The hot particles associated with particulate matter sampled at the sea surface, and hence available for transfer into marine aerosols, have a mean particle diameter suitable for entry into the lung of man. In relation to potential health hazards to man the levels of currently released radionuclide effluents may be of more concern than those retained in older sediments where hot particles are less abundant and a diffuse distribution of α activity is more common, possibly reflecting redistribution as a result of geochemical recycling. Apart from the release of transuranics in effluent into the sea there is also a need to consider releases from the BNFL stacks and from the uncovered uranium fuel rod cooling ponds. Cambray and Eakins¹⁸ have shown that Pu is present in soils up to ~10 km from the Cumbrian coast and is either derived from direct emissions from Windscale or indirectly from the sea. Gorham¹⁹ has shown that within the region, up to a distance of at least 51 km from the coast, almost all the sodium and chloride ions in tarns (altitude 15–275 m) is derived from sea spray. Cambray and Eakins noted a correlation between Na and Pu and favour the sea surface as a major source for the Pu. Apart from the transport of conservative species of Pu and Am in droplets of seawater in marine aerosols the possibility exists for entrainment of the small sea surface hot particles which I have also found in soils, lichens and one ombrotrophic peat inland from the coast. So far their distribution has not been evaluated because of difficulties in developing an acceptable programme of valid statistical sampling of such diffuse distributions.

Using conventional air sampling methods Pattenden *et al.*²⁰ note that Pu isotopes, ²⁴¹Am and ¹³⁷Cs are present in airborne and deposited material in west Cumbria in excess of the quantities expected for nuclear weapon tests. Excess actinides are considered to be derived from sea-spray and surf, but based on measurements of ¹³⁷Cs this cannot be accounted for by transfer of bulk seawater in the form of droplets. They note that conventional sampling methods are not efficient for samples with a mean particle diameter of >20 μ m because the inertia of large particles limits an ability to follow a moving air-stream. Using the CR-39 method I have observed that while the highest specific activity of resuspended air and sand particles is usually associated with those of a mean particle diameter of <20 μ m, there is an abundance of uniformly-labelled large particles which consist of coated sand grains (mean particle diameter 100–300 μ m) and ferruginous flocs¹⁵. In an examination of ~1,000 particles from the Esk region (resuspended silts, rocks, lichens) only one ~500 μ m particle was found to have a specific activity comparable with the sea-surface hot spots. The contribution of diffuse labelled large particles to sea-surface type hot particles is being investigated and is likely to be significant. CR-39 is capable of identifying the distribution of the hot particles on foliage and the surface of soils distant from Windscale and such a programme has been implemented. Because of the considerable uncertainties in even measuring maximum permissible levels of Pu in the human lung it would be useful to determine the distribution and composition of these hot particles in Cumbria and adjacent areas and evaluate their importance to man.

I thank R. J. Clifton and P. J. Moore for useful discussions. The Institute for Marine Environmental Research is a component of the Natural Environment Research Council.

Received 24 September 1980; accepted 10 February 1981.

1. Hamilton, E. I. *Symp. on Detection of Radiation in Environmental and Biological Materials* (CEGB, London, 1978).
2. Hamilton, E. I. *Mar. Ecol. Prog. Ser.* **2**, 61–73 (1980).
3. Hamilton, E. I. & Clifton, R. J. *Mar. Ecol. Prog. Ser.* **3**, 267–277 (1980).
4. Hamilton, E. I. & Clifton, R. J. *Int. J. Appl. Isotopes Radiat.* (in the press).
5. Cartwright, R. G., Shirk, E. K. & Price, P. B. *Nucl. Instrum. Meth.* **153**, 457–460 (1978).
6. Evans, R. D. A. *Prog. Rep. AT (30-1)-952-MIT-952-4*, 132–162 (Radioactivity Center, MIT 1967).
7. Nair, S. U.K. *C.E.G.B. Rep. RD/B/N4349*, 1–34 (1978).
8. Hetherington, J. A. & Jefferies, D. F. *Neth. J. Sea. Res.* **8**, 319 (1974).
9. Pentreath, R. J., Lovett, M. B., Harvey, B. R. & Ibbett, R. D. *Biological Implications of Radionuclides Released from Nuclear Industries Vol. II*. (IAEA, Vienna, 1979).
10. Baxter, M. S., McKinley, I. G. M., Mackenzie, A. B. & Jack, W. *Mar. Pollut. Bull.* **10**, 116–120 (1979).

11. Aston, S. R. & Stanner, D. A. *Nature* **289**, 581–582 (1981).
12. Clifton, R. J. & Hamilton, E. I. *Est. Coast. Shelf Sci.* (submitted).
13. Hodge, V. F., Koide, M. & Goldberg, E. D. *Nature* **227**, 206–209 (1979).
14. A. Rep. *Radiation Discharges and Monitoring of the Environment* 1978 (British Nuclear Fuels, London 1979).
15. Hamilton, E. I., Clifton, R. J. & Stevens, H. (in preparation).
16. Hampson, B. L. & Tennant, D. *Analyst* **98**, 873–885 (1973).
17. Hetherington, J. A., Jefferies, D. F. & Lovett, M. B. In *Impacts of Nuclear Releases into the Aquatic Environment*, 210 (IAEA, Vienna, 1975).
18. Cambray, R. S. *UKAEA AERE Rep. R-9807*, HL80/2061 (C10) 1–15 (HMSO, London, 1980).
19. Gorham, E. *Phil. Trans. R. Soc. B*, 147–178 (1958).
20. Pattenden, N. J., Cambray, R. S., Playford, K., Eakins, J. D. & Fisher, E. M. R. *UKAEA AERE Rep. R-9857*, HL80/2746 (C10) 1–22 (HMSO, London, 1980).
21. Hunt, G. J. *Aquatic. Envir. Mon. Rep. MAFF—FRL No. 3*, 1–36 (1979).

Specific acyclic isoprenoids as biological markers of methanogenic bacteria in marine sediments

S. C. Brassell, A. M. K. Wardroper, I. D. Thomson, J. R. Maxwell & G. Eglinton

Organic Geochemistry Unit, University of Bristol, School of Chemistry, Cantock's Close, Bristol BS8 1TS, UK

The widespread occurrence of extended hopanoids in sediments and petroleum illustrates the importance of bacterial lipid contributions to geological materials¹. In archaeobacteria, however, hopanoids are absent²; their role as structural components of biomembranes is fulfilled by acyclic isoprenoids^{1–5}. Recent studies^{2–6} of the lipid constituents of archaeobacteria have greatly extended the range of acyclic isoprenoid skeletons known in organisms (Fig. 1). In particular, isoprenoids with head-to-head linkages have been identified^{3–5}, and such compounds (for example, 3,7,11,15,18,22,26,30-octamethyldotriacontane, I) have been recognized in petroleum^{7,8} and as degradation products of Messel shale kerogen^{9–11}. Here we report the first recognition of 2,6,10,15,19-pentamethyleicosane (II), a known component of methanogens^{5,6}, in marine sediments of Recent to Cretaceous age (Table 1) and suggest that it and certain other acyclic isoprenoids may be used as biological markers for methanogens.

Until recently, the geologically occurring isoprenoid alkane, squalane (III), first characterized in a Nigerian oil¹², was generally believed to originate from geochemical reduction of squalene (IV). The discovery of squalane in several immature marine sediments^{13–17}, including Recent sediments^{13,14}, suggested that such a reduction must occur rapidly in the upper layers of sediment and therefore be microbially mediated. An alternative explanation is that squalane represents a direct input from organisms and the recent recognition of squalane in archaeobacteria^{5,6} supports this proposition. If, in general, archaeobacteria are important contributors of lipids to sediments, one would expect to find other compounds uniquely biosynthesized by these organisms in geological materials. One such compound seems to be 2,6,10,15,19-pentamethyleicosane, which has been found in methanogens^{5,6} but not in the other types of archaeobacteria, or in marine phytoplankton, such as diatoms, dinoflagellates, silicoflagellates and coccolithophores, which are major sources of sediment lipids. In addition, it has not been detected in zooplankton, although these organisms frequently contain another acyclic isoprenoid alkane, pristane, derived from chlorophyll¹⁸.

2,6,10,15,19-Pentamethyleicosane (II), squalane (III) and lycopane (V) were identified in the sediments by comparison of their mass spectra, obtained by computerized gas chromatography–mass spectrometric¹⁹ analysis of hydrocarbon fractions, with literature and/or reference spectra. Each sediment was extracted using ultrasonication; the total lipid extract was separated by TLC yielding a hydrocarbon fraction which, in some cases, was further separated by urea adduction. The mass

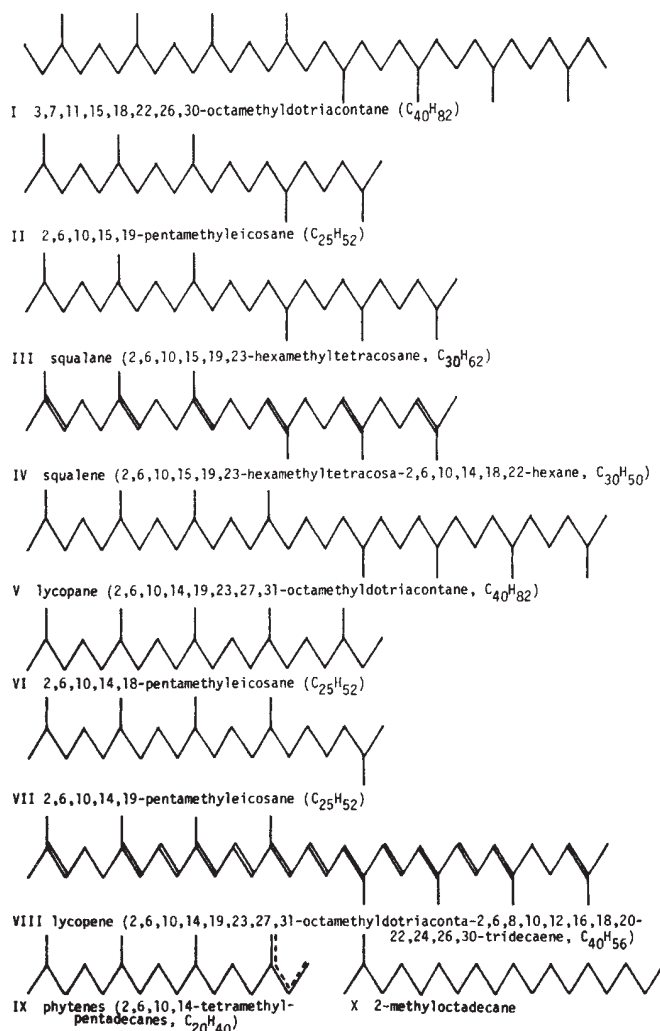
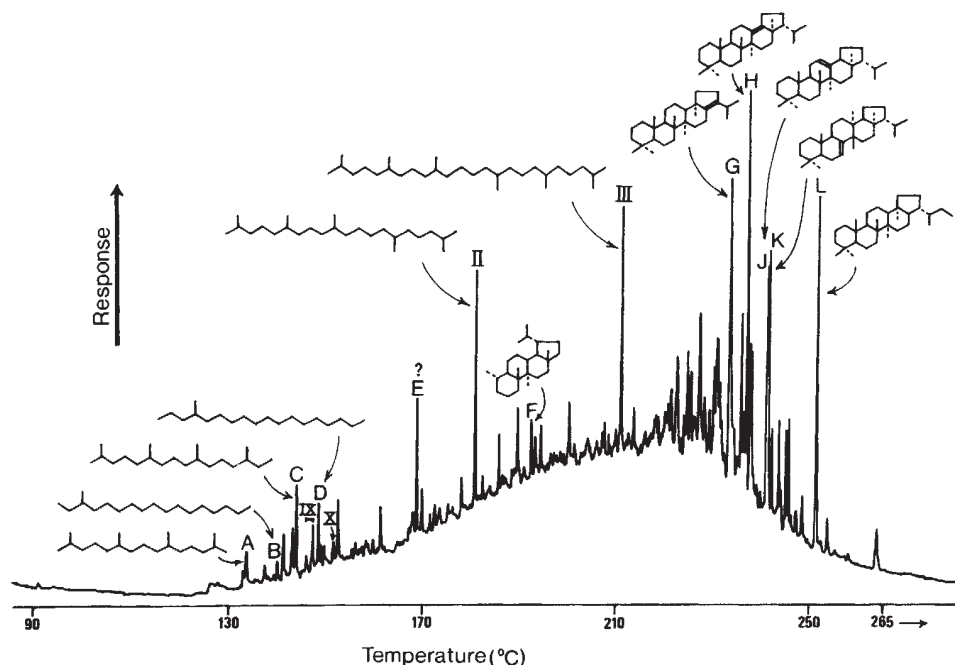


Fig. 1 Structures of acyclic hydrocarbons.

Fig. 2 Gas chromatogram (GC) of the branched/cyclic hydrocarbons of a Pleistocene diatomaceous clay/silty claystone (DSDP Leg 57, sample 57-440B-8-4; Table 1) from the Japan Trench. Peaks: A, pristane; B, 3-methylheptadecane; C, phytane; IX, phytanes; D, 4-methyloctadecane; X, 2-methyloctadecane; E, $C_{25}H_{44}$ bicyclic alkadiene of unknown structure^{44,45}; II, 2,6,10,15,19-pentamethyleicosane; F, tetracyclic alkane related to lupane^{46,47}; III, squalane; G, hop-17(21)-ene; H, neohop-13(18)-ene; J, neohop-12-ene; K, fern-7-ene; L, 17 β H,21 β H-homohopane. GC conditions: 20 m OV-1 wall-coated glass capillary, programmed from 50 to 265 °C at 4 °C per min after splitless injection at ambient temperature. A prominent feature of the GC trace is the apparent 'hump' which in crude oils and polluted sediments would be attributed to a background of thermally matured hydrocarbons. In this sample, however, the concentrations of other individual hydrocarbons (for example, steranes and hopanes) commonly associated with the 'hump' is small; hence, it is uncertain whether it represents highly biodegraded thermally matured or natural inputs or, alternatively, shipboard contamination with drilling lubricant.



spectrum of 2,6,10,15,19-pentamethyleicosane (II) is distinct from those of VI (2,6,10,14,18-pentamethyleicosane) and VII (2,6,10,14,19-pentamethyleicosane) for several reasons^{6,16,17}, particularly the absence of m/z 253 in II. The assignments of squalane (III) and lycopane (V) were corroborated in a number of cases by gas chromatographic co-injection with authentic standards. Components were quantified from their gas chromatographic responses.

The recognition of 2,6,10,15,19-pentamethyleicosane (II) in three marine environments where methanogens have been recognized and/or high concentrations of methane have been found (Table 1) suggests the working hypothesis that this alkane marks a contribution of lipids from methanogens to marine sediments. An isomer, 2,6,10,14,18-pentamethyleicosane (VI), recently found in thermoacidophilic bacteria⁶, has previously been reported as a constituent of various shales²⁰, oils^{7,8,21-23} and sediments²⁴. The occurrence of squalane (III) may also reflect methanogenic activity in these marine sediments (Table 1), although this compound has been found in both thermoacidophilic and methanogenic bacteria⁶. However, significant inputs to marine sediments from thermoacidophilic and halophilic bacteria are improbable, although such archaeobacteria may be important sources of lipids in other environments, for example in the Dead Sea²⁵. Lycopane (V) is another isoprenoid alkane found in significant quantities (see Table 1) in Recent^{14,16,17} and ancient²⁶ marine sediments. It could be formed by diagenetic reduction (perhaps microbial) of lycopene (VIII)^{14,16,17} but, by analogy with squalane, a direct bacterial origin is also possible. Indeed, the Recent sediments in which lycopane has been reported are all environments where methanogenic activity has been documented (Table 1).

In older and more deeply buried sediments, after methanogenic biological activity has ceased, the presence of methane and the lipid signatures may attest its former existence. Hence, in the Pleistocene to Miocene sediments of the Cariaco and Japan Trenches, the presence of methane in high concentrations²⁷⁻²⁹ and the recognition of the isoprenoid alkanes II and III (Table 1) indicate past biological methanogenesis in these sediments. In more mature sediments the presence of methane alone cannot be used as an indicator of former methanogenic biological activity. For example, the methane in the Moroccan Basin turbidites (Table 1) is thought to be migrated rather than

Table 1 Concentrations of 2,6,10,15,19-pentamethyleicosane (II), squalene (III) and lycopane (V) in selected marine sediments

Location (lat, long)	Sub-bottom depth(m)	Geological age (Myr BP)	Dominant lithology	Concentrations (ng per g dry sediment weight)			
Sample codes				II	III	V	CH ₄
Walvis Bay (22°35'S, 13°45'E) 1/75 (0–0.1)	0.05	Holocene (<0.01)	Diatomaceous ooze	250	ND	180	High*
Gulf of California (27°15'N, 111°31'W) 64-481-2-2	7.5	Holocene (<0.01)	Diatomaceous ooze	16	23	68	High†
Cariaco Trench (10°43'N, 65°11'W)							
15-147B-1-2	3	Holocene (<0.01)	Calcareous clay	1.1	0.7	14	} High‡
15-147B-5-5	48	Pleistocene (~0.1)	Calcareous clay	+	+	+	
15-147C-3-3	138	Pleistocene (~0.25)	Calcareous clay	6.7	5.3	57	
Japan Trench (39°44'N, 143°48'E)							
56-435-13-2	115	M. Pleistocene (~3)	Diatomaceous clay	0.8	0.4	ND	446§
(39°56'N, 145°33'E) 56-436-11-4	99	U. Pliocene (~2)	Diatomaceous vitric clay	0.3	0.6	ND	5
(39°44'N, 143°56'E)							
57-440-5-6	42.5	U. Pleistocene (~0.5)	Diatomaceous silty clay	0.3	0.8	ND	115
57-440A-7-6	139	L. Pleistocene (~1)	Diatomaceous silty clay	1.5	2.8	ND	842
57-440B-3-5	166	L. Pleistocene (~1)	Diatomaceous clay	1.9	3.1	ND	803
57-440B-8-4	212	L. Pleistocene (~1)	Diatomaceous clay/silty claystone	1.7	1.7	ND	1596
57-440B-23-4	354	U. Pliocene (~2)	Diatomaceous claystone/silty claystone	0.5	0.4	ND	151
57-440B-39-3	504	L. Pliocene (~4)	Diatomaceous claystone	2.8	1.8	ND	215
57-440B-68-2	778	U. Miocene (~6)	Silty claystone	1.4	1.7	ND	289
Blake-Bahama Basin (28°14'N, 75°37'W)							
44-391A-13-0	525	L. Miocene (~20)	Siliceous mudstone	+	ND	ND	} NA
44-391C-6-3	691	U. Albian (~100)	Claystone	+	+	ND	
44-391C-12-4	959	Aptian (~110)	Calcareous claystone	ND	ND	ND	
Moroccan Basin (32°50'N, 10°48'W)							
50-416A-9-2	1179	Hauterivian (~125)	Turbidite	0.4	0.5	ND	11.4
50-416A-14-4	1228	Valanginian (~130)	Turbidite	0.3	0.5	ND	ND
50-416A-18-2	1266	Valanginian (~130)	Turbidite	0.1	Tr	ND	ND
50-416A-23-3	1313	Valanginian (~130)	Turbidite	Tr	0.2	ND	14.9
50-416A-28-4	1362	Valanginian (~130)	Turbidite	Tr	0.1	ND	ND

Detailed analyses of the lipid distributions of II, III and V are reported elsewhere^{15–17,36–38}. ND, not detected; NA, not analysed; +, minor amounts not quantified; Tr, trace (<0.1 ng per g) amounts of component present.

*CH₄ found (90–2,926 and 10–91 ng per g) in subsections of sediment cores up to 2.3-m depth from nearby locations (22°47'S, 14°26'E and 23°14'S, 14°01'E, respectively) and attributed to methanogenic bacterial activity²⁷.

†CH₄ occurs in Site 481 sediments³⁹ and is abundant (2,190–7,520 ng per g) in the subsurface sediments (2.51–3.45 m) collected about 8 km to the north-east (27°23'N, 111°27'W), with $\delta^{13}\text{C}$ values (–68.8 to –73.5) indicating its biogenic origin⁴⁰. Methanogenic bacterial activity has been recognized in an overlying sample (64-481-1-2) from 2.9-m depth⁴¹ and shallow sediments (0.5–1.7 m) from a nearby site (27°09'N, 111°08'W)⁴².

‡CH₄ detected in gas pockets to 175-m depth at Site (147 (ref. 28) and found in high concentrations (>200 ng per g) in other shallow sediments (0–1.6 m) from the area²⁷. Methanogenic bacteria have been isolated from nearby surface sediments within the western basin of the Trench (10°37'N, 65°26'W)⁴³. Lycopane has also been recognized in shallow sediments (0–5.5 m) from the eastern basin (10°32'N, 64°39'W)¹⁴.

§Data for adjacent samples from ref. 29.

||Data for adjacent samples from ref. 30.

indigenous gas³⁰. Hence, the presence of the acyclic isoprenoid alkanes provides a better geological record of biological methanogenesis than the presence of methane.

In the Holocene sediments the differences in the relative proportions of II, III and V may result from dissimilar methanogen populations, as methanogen lipid compositions are species dependent^{5,6}. In addition, the variability in their concentrations in these sediments may reflect the intensity of biological methanogenesis. The deeper and older sediments tend to contain lower concentrations than do the Holocene sediments (Table 1), partly because other microorganisms may have degraded the methanogenic lipids. Other factors may be involved, for example, the known flux of terrestrial material in the Japan Trench sediments and Moroccan Basin turbidites may have diluted the methanogenic lipid signature³¹. Such factors, however, tend to restrict rather than invalidate the use of acyclic isoprenoid alkanes as biological markers of methanogens. Although the Cretaceous samples contain only low concentrations of II and III, the mere presence of these compounds is presumptive evidence for methanogenesis at the time of deposition. Methanogenic bacteria seem, therefore, to be important contributors of lipids to deep-sea sediments at present and in the geological past, presumably as a result of their activity at appropriate E_h (redox potential) zones in the sub-bottom sediments. The significance of archaeobacterial contri-

butions to petroleum has already been demonstrated^{7,8}, but present evidence suggests that specific acyclic isoprenoid alkanes may also be environmental indicators. That they are important constituents of sediments can be seen from their relative abundances in branched/cyclic (urea non-adducted) hydrocarbon fractions (Fig. 2). Future investigations should include stereochemical determination of these acyclic isoprenoid alkanes in sediment extracts, as their configurations will reflect their origin and diagenetic history^{32,33}, together with studies of the stereochemistry of these compounds in methanogenic bacteria.

Other isoprenoid hydrocarbons may also arise from methanogens. Squalene (IV) and phytanes (IX) have been found in young marine sediments^{16,17}, but could originate from other sources—for example, phytanes may be products of phytol or dihydrophytol diagenesis³⁴. Thus, these compounds cannot generally be regarded as definitive indicators of archaeobacterial inputs to sediments. An additional group of acyclic hydrocarbons biosynthesized by archaeobacteria⁶ is branched alkanes (for example, 2-methyloctadecane, X); these compounds also occur in deep-sea sediments¹⁷ (Fig. 2). Indeed, most of the prominent branched/cyclic hydrocarbons of the Japan Trench sediments (Fig. 2) can be ascribed to bacterial origins. In view of their recent recognition in a bacterium³⁵, the presence of hop-17(21)-ene (G), neohop-13(18)-ene (H) and fern-7-ene (K) in

immature sediments probably reflects autochthonous inputs from bacteria rather than terrestrial inputs from ferns¹⁵⁻¹⁷. It seems likely, by analogy, that neohop-12-ene (J) is also of bacterial origin.

Thus, there is growing evidence in support of major bacterial lipid inputs to marine sediments. The presence of components which seem to be biosynthesized uniquely by methanogens, such as 2,6,10,15,19-pentamethyleicosane, suggests that these bacteria are important contributors of alkanes to marine sediments.

We thank the NERC (GR3/2951 and GR3/3758) and NASA (subcontract from NGL-05-003-003) for funding. A.M.K.W. and I.D.T. acknowledge studentships from the SRC, S.C.B. from the NERC. We thank Dr R. S. Oremland for information about methanogens in Gulf of California sediments, Mrs A.P. Gowar for technical assistance, and colleagues at Bristol for access to extracts^{15,19}. Sediment samples were provided by the Deep Sea Drilling Project (JOIDES), with the exception of the Walvis Bay sediment, which was kindly supplied by Dr R. J. Morris.

Received 11 August 1980; accepted 4 February 1981.

1. Ourisson, G., Albrecht, P. & Rohmer, M. *Pure appl. Chem.* **61**, 709-729 (1979).
2. Kates, M. in *Either Lipids, Chemistry and Biology* (ed. Synder, F. S.) 351-398 (Academic, New York, 1972).
3. Tornabene, T. G. & Langworthy, T. A. *Science* **203**, 51-53 (1979).
4. De Rosa, M., De Rosa, S., Gambacorta, A. & Bu'Lock, J. D. *Chem. Commun.* 514-515 (1977).
5. Tornabene, T. G., Langworthy, T. A., Holzer, G. & Oro, J. *J. molec. Evol.* **13**, 73-83 (1979).
6. Holzer, G., Oro, J. & Tornabene, T. G. *J. Chromatogr.* **186**, 795-809 (1979).
7. Moldowan, J. M. & Seifert, W. K. *Science* **204**, 169-171 (1979).
8. Albaiges, J. in *Advances in Organic Geochemistry 1979* (eds Douglas, A. G. & Maxwell, J. R.) 19-28 (Pergamon, Oxford, 1980).
9. Michaelis, W. & Albrecht, P. *Naturwissenschaften* **66**, 420-422 (1979).
10. Chappe, B., Michaelis, W. & Albrecht, P. in *Advances in Organic Geochemistry 1979* (eds Douglas, A. G. & Maxwell, J. R.) 265-274 (Pergamon, Oxford, 1980).
11. Chappe, B., Michaelis, W., Albrecht, P. & Ourisson, G. *Naturwissenschaften* **66**, 522-523 (1979).
12. Gardner, P. M. & Whitehead, E. V. *Geochim. cosmochim. Acta* **36**, 259-263 (1972).
13. Dastillung, M., Albrecht, P., Tissier, M. J. in *Géochimie Organique des Sédiments Marins Profonds. Orgon I; Mer de Norvège*, 209-228 (CNRS, Paris, 1977).
14. Dastillung, M. & Corbet, M. in *Géochimie Organique des Sédiments Marins Profonds. Orgon II; Atlantique-N. E. Brésil*, 293-323 (CNRS, Paris, 1978).
15. Brassell, S. C. et al. *Init. Rep. DSDP Leg. 56/57*, Part 2, 1367-1390 (1980); in *Advances in Organic Geochemistry 1979* (eds Douglas, A. G. & Maxwell, J. R.) 375-392 (Pergamon, Oxford, 1980).
16. Wardroper, A. M. K. thesis, Univ. Bristol (1979).
17. Brassell, S. C. thesis, Univ. Bristol (1980).
18. Blumer, M., Mullin, M. M. & Thomas, D. W. *Science* **140**, 974 (1963).
19. Brassell, S. C., Gowar, A. P. & Eglinton, G. in *Advances in Organic Geochemistry 1979* (eds Douglas, A. G. & Maxwell, J. R.) 421-426 (Pergamon, Oxford, 1980).
20. Spyckerelle, C., Arpino, P. & Ourisson, G. *Tetrahedron* **28**, 5703-5713 (1972).
21. Albaiges, J., Borbon, J. & Salagre, P. *Tetrahedron Lett.* 595-598 (1978).
22. Han, J. & Calvin, M. *Geochim. cosmochim. Acta* **33**, 739-742 (1969).
23. Haug, P. & Curry, D. J. *Geochim. cosmochim. Acta* **38**, 601-610 (1974).
24. Waples, D. W., Haug, P. & Welte, D. H. *Geochim. cosmochim. Acta* **38**, 381-387 (1974).
25. Kaplan, I. R. & Baedeker, M. J. *Israel J. Chem.* **8**, 529-533 (1970).
26. Comet, P. A. et al. *Init. Rep. DSDP Leg. 62* (in the press).
27. Whelan, J. K., Hunt, J. M., & Berman, J. *Geochim. cosmochim. Acta* **44**, 1767-1785 (1980).
28. Hammond, D. E., Horowitz, R. M., Broecker, W. S. & Bopp, R. *Init. Rep. DSDP Leg. 20*, 765-771 (1972).
29. Whelan, J. K. & Hunt, J. M. *Init. Rep. DSDP Leg. 56/57*, Part 2, 1349-1365 (1980).
30. Whelan, J. K. & Hunt, J. M. *Init. Rep. DSDP Leg. 50*, 623-624 (1980).
31. Von Huene, R. et al. *Geotimes* **23**(4), 16-21 (1978).
32. Cox, R. E., Maxwell, J. R., Ackman, R. G. & Hooper, S. N. *Can. J. Biochem.* **50**, 1238-1241 (1972).
33. Patience, R. L., Rowland, S. J. & Maxwell, J. R. *Geochim. cosmochim. Acta* **42**, 1871-1875 (1978).
34. Didyk, B. M., Simoneit, B. R. T., Brassell, S. C. & Eglinton, G. *Nature* **272**, 216-222 (1978) and refs therein.
35. Howard, D. L. thesis, Univ. California, Los Angeles (1980).
36. Cardoso, J. N. et al. *Init. Rep. DSDP Leg. 44*, 965-976 (1979).
37. Brassell, S. C. et al. *Init. Rep. DSDP Leg. 50*, 647-664 (1980).
38. Thomson, I. D., Brassell, S. C., Eglinton, G. & Maxwell, J. R. *Init. Rep. DSDP Leg. 64*, (in the press).
39. Curran, J. R. et al. *Geotimes* **24** (7), 18-20 (1979).
40. Simoneit, B. R. T., Mazurek, M. A., Brenner, S., Crisp, P. T. & Kaplan, I. R. *Deep Sea Res.* **26A**, 879-891 (1979).
41. Oremland, R. S., Culbertson, C. & Simoneit, B. R. T. *Init. Rep. DSDP Leg. 64*, (in the press).
42. Belyaev, S. S., Lein, A. Yu. & Ivanov, M. V. in *Biogeochemistry of Ancient and Modern Environments* (eds Trudinger, P. A., Walter, M. R. & Ralph, B. J.) 235-242 (Australian Academy of Science, Canberra, 1980).
43. Romesser, J. A., Wolfe, R. S., Mayer, F., Spiess, E. & Walter-Mauruschat, A. *Arch. Microbiol.* **121**, 147-153 (1979).
44. Farrington, J. W., Frew, N. M., Gswend, P. M. & Tripp, B. W. *Estuarine Coastal mar. Sci.* **5**, 793-808 (1977).
45. Boehm, P. D. & Quinn, J. G. *Estuarine Coastal. mar. Sci.* **6**, 471-494 (1978).
46. Kimble, B. J. thesis, Univ. Bristol (1972).
47. Corbet, B., Albrecht, P. & Ourisson, G. *J. Am. chem. Soc.* **102**, 1171-1173 (1980).

Anomalous ¹³C depletion in early Precambrian graphites from Superior Province, Canada

M. Schoell

Bundesanstalt für Geowissenschaften und Rohstoffe, Stilleweg 2, 3000 Hannover 51, FRG

F.-W. Wellmer*

Metallgesellschaft AG Reuterweg 14, 6000 Frankfurt a.M., FRG

The carbon isotope composition of organic carbon is potentially, at least, a guide to the conditions in which and to the organisms by which it was formed. The isotopic composition of most Phanerozoic kerogens, measured as $\delta^{13}\text{C}$ values, lies within the range -20‰ to -30‰. Most (~65%) samples of Precambrian organic carbon, now represented by graphite and other forms of reduced carbon, have isotopic ratios within the same range, but in others the concentration of ¹³C is significantly depleted by as much as 10‰. The causes of this depletion have an obvious bearing on the conditions of life in the Precambrian. Here we report the results of our measurements of samples of graphite from the Archean of the Superior Province of Canada, one of which being the most depleted in ¹³C ever found in the Precambrian. We explain this ¹³C depletion by organisms in some areas living on CO₂ oxidized from CH₄ that has been formed by methane-producing bacteria in Precambrian stratified seas.

Some 35% of the Precambrian graphite samples measured have ¹³C concentrations which lie outside the Phanerozoic range. Enrichment has usually been explained by the preferential loss of ¹²C during metamorphism¹⁻⁴, but depletion is harder to account for. It has been suggested that depletion may be a consequence of higher CO₂ partial pressure in the Precambrian^{5,6} or may have come about because of the predominance of lipids depleted in ¹³C in Precambrian organic matter⁷.

All ¹³C/¹²C ratios of graphite samples from the Archean of the Superior Province studied so far were within the range of Phanerozoic kerogens⁸. They were from graphites associated with sulphidic and oxidic iron formations and their transition zone. Iron formations are commonly intercalated between volcanoclastic sediments, less common between volcanic flows, and closely associated with cherts, carbonaceous argillites and graywackes. Iron formation horizons can often be followed for tens of kilometres indicating extensive sedimentary basins; the Michipicoten area of northwestern Ontario is a typical example. The isotope ratios of graphites from iron formation horizons of this area all lie in the range -20 to -28‰ (ref. 8). This range also covers the only sample of this report collected in a similar sulphidic iron formation environment of the Savant Lake-Sturgeon Lake area in northwestern Ontario (102 PQ, Table 1).

The other nine samples on which we report were collected from drill holes in a different type of sulphidic graphitic sediments which occur within the volcanic piles themselves. These sediments consist of thin layers of graphitic sulphidic tuffs or interflow sediments with strike lengths of the order of only 100-1,000 m. In contrast to the large sedimentary basins with extensive iron formations these short graphitic horizons are imbedded in local depressions the size of which is indicated by the strike lengths of the graphite horizons derived from airborne and ground geophysics. One of these horizons extends over a distance of 1,400 m but all others stretch only a few hundred metres indicating rather small pools.

* Present address: Bundesminister für Wirtschaft Villemombler Straße, 5300 Bonn 1, FRG.

Table 1 Field data and analytical results

Samples	Lat. Long.	Location	S* L† T‡	$\delta^{13}\text{C}_{\text{PDB}}$ (‰)	Geological environment	
			(m)		Footwall	Hangingwall
102 PQ	50°23'N 90°34'W	Conant Township NW Ontario Wabigoon Belt	0 8,000 4	-25.9	Rhyolite tuffs	Intermediate to mafic flows and tuffs
U 6	50°23'N 90°34'W	Conant Township NW Ontario Wabigoon Belt	800 1,400 2.4	-22.5	Dacitic tuffs and shales	Dacitic tuffs and shales
A 3/1§	47°08'N 84°14'W	Township 26, R Range 13 (Olsen) Ontario, Batchawana Wawa Belt	1,200 250 0.9	-26.9	Mafic to intermediate tuffs	Mafic to intermediate tuffs
A 3/2§	47°08'N 84°14'W	Township 26, R Range 13 (Olsen) Ontario, Batchawana Wawa Belt	1,200 250 0.9	-26.9	Mafic to intermediate tuffs	Mafic to intermediate tuffs
S 2	49°16'N 89°42'W	Heaven Lake NW Ontario Wabigoon Belt	800 210 15	-24.6	Cherty sediments	Cherty sediments and massive sulphides
K 1	49°59'N 87°42'W	Onaman Lake area NW Ontario Wabigoon Belt	400 270 3.1	-21.0	Mafic fragmentals	Siltstone and mafic tuffs
K 10	49°59'N 87°42'W	Onaman Lake area NW Ontario Wabigoon Belt	400 270 1.7	-25.2	Mafic fragmentals	Siltstone and mafic tuffs
K 2¶	49°59'N 87°42'W	Onaman Lake area NW Ontario Wabigoon Belt	1,100 150 4.7	-35.5	Tuffaceous shales	Tuffaceous shales
K 6	49°50'N 78°41'W	Sandra Township NW Ontario Wabigoon Belt	900 275 3.3	-38.3	Rhyodacite	Andesite
X-1	48°50'N 78°41'W	Languedoc Township Quebec, Abitibi Belt	700 360 5.0	-43.8	Interbedded mafic to intermediate tuffs and shales	Mafic flows

The samples were crushed and treated with diluted acid to remove possible carbonate material, and then burned at temperatures of 900 °C in a stream of pure dried oxygen using a copper oxide catalyst. The resulting CO₂ was analysed for its carbon isotopic composition with a triple collecting mass-spectrometer (VG 903). Usual corrections were applied and the isotopic composition is reported as δ -value versus the PDB-standard. All samples were processed twice. The reproducibility as determined from double measurements is $\pm 0.2\%$. The NBS graphite standard has been determined at $-28.2 \pm 0.1\%$. (δ -values are defined as relative deviations from the PDB standard where $\delta^{13}\text{C} = (R_{\text{sample}}/R_{\text{standard}} - 1) \times 1,000\%$ where R is the isotope ratio $^{13}\text{C}/^{12}\text{C}$.)

* Stratigraphic distance to top of volcanic pile.

† Length of graphite horizons.

‡ Thickness of graphite horizons.

§ Two graphite horizons in the same drill hole separated by a 15 m stratigraphic section.

|| K 1 and K 10 same graphite horizon 60 m apart along strike.

¶ Horizon of K 2, 700 m stratigraphically lower than K 1 and K 10.

The volcanic and sedimentary host rocks of the graphite horizons have been metamorphosed in greenschist facies conditions. Typical mineral assemblages in the volcanics are chlorite, actinolite, epidote, albite, quartz and sericite. Samples A3/1, A3/2 and X-1 are from the Abitibi-Wawa volcanic plutonic belt for which radiometric data indicate an age of $\sim 2,750$ Myr; all other samples are from the 2,800-Myr old Wabigoon volcanic plutonic belt^{9,10}. Our graphite samples come from bands and layers intricately intercalated with volcanoclastic sediments, indicative of their indigenous nature. Microscopic examination shows that graphites form flakes which are commonly oriented parallel to the schistosity; optically the flakes display a clearly visible pleochroism which proves the graphitic nature of the carbonaceous matter.

Table 1 gives the results of our isotope analyses and some field data, obtained from exploration drilling. The geological position of the graphite horizons is estimated from their stratigraphic distance to the top of the volcanic pile which varies between 1,200 and 0 m, mostly being 400–900 m.

Seven of the 10 samples analysed revealed $\delta^{13}\text{C}$ values between -21 and -27% —the normal range within the spread of data of many other Precambrian samples (Fig. 1). Three samples show significant depletion in the heavy isotope with $\delta^{13}\text{C}$ values between -35% and as low as -43.8% ; this value is among the lowest ^{13}C concentration ever reported for Precambrian reduced carbon. The field data indicate a common geological setting even for those localities which yielded the anomalously low ^{13}C graphite data. Sample K2 from the Onaman Lake area in northwestern Ontario with an extremely low $\delta^{13}\text{C}$ value was taken in the same volcanic pile only 700 m away from two normal value samples (K1 and K10).

Anomalous ^{13}C depletion ($\delta^{13}\text{C}$ less than -30%) in Precambrian organic matter has been reported previously^{4,5,12,14,16} (Fig. 1). But some of the values shown in Fig. 1 are not indicative of the primary source. For example, $\delta^{13}\text{C}$ values of -37.01 and -39.1% respectively have been reported for the so-called anthraxolites⁴. Stringers of visible carbon in limestones were also found to be depleted in ^{13}C ($\delta^{13}\text{C} \approx -37$ to -41%) as

compared with disseminated carbon which had 'normal' values around -26 to -33% (ref. 16). Secondary effects must have been operative to form both the anthraxolites and the stringers which forbids any conclusion as to their original isotopic composition. Such secondary effects are ruled out for our graphite samples (data set 1 in Fig. 1). They are undoubtedly of syngenetic nature which is indicated by the above described primary sedimentary features.

Thus the anomalous $^{13}\text{C}/^{12}\text{C}$ isotope ratios indicate special environmental conditions which must have prevailed in some of these small depressions. These special conditions are assumed to have prompted the organic species to incorporate the ^{12}C -isotope more readily.

We discarded the theory that in some pools specific assemblages of organisms had different isotope fractionations when assimilating carbon. Also, we consider it unlikely that various differing metabolic systems were involved.

The change of the isotopic composition of the carbon source is therefore considered in the following as the process which may easily explain such a large change in organic $^{13}\text{C}/^{12}\text{C}$ ratios as we observed or even larger changes. In all cases the formation of graphite is closely interrelated with the occurrence of sulphidic sediments indicating areas of hydrothermal activity. The discharge of hydrothermal brines into morphological depressions is likely to cause a stratification of waters as described for the Red Sea brine pools, the Lake Kivu or more recently from below the Gulf of Mexico¹⁷⁻²⁰. The brines prevent vertical mixing and thereby act as traps for organic debris. Methane can form in such environments and may have become trapped in the stratified waters. The formation of methane by bacteria is the only geochemical process known to bring about appreciable ^{13}C -depletion (typical $\delta^{13}\text{C}$ values are around -60 to -80%). Methane may become oxidized to CO_2 when appreciable residence times within the stratified waters allow CH_4 oxidizing organisms to grow²⁴. The CO_2 from methane oxidation is even more depleted in ^{13}C than the methane²¹ and its ^{13}C concentration can range from -70 to -90% . Such ^{13}C -depleted CO_2 would then be admixed to CO_2 which may have been derived

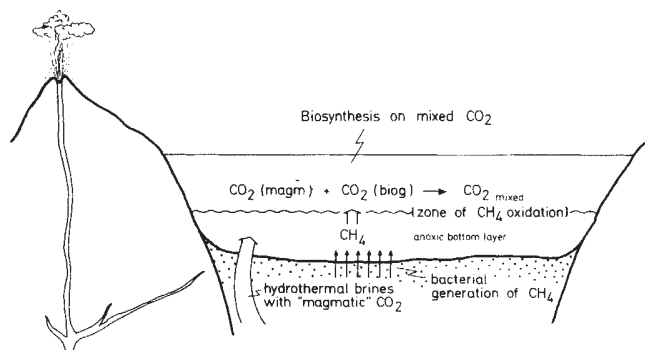


Fig. 2 Hypothetical explanation of the ^{13}C depletion in Precambrian organic carbon. Stratified waters are traps for methane which is bacterially produced in the stratified waters. Oxidation in the sediment or at water interfaces would generate ^{13}C depleted CO_2 as a carbon source for metabolizing eukaryotic biota.

from hydrothermal solutions or from air CO_2 to form a carbon reservoir, depleted in ^{13}C (Fig. 2). Oxidation of CH_4 to CO_2 does not necessarily need free atmospheric oxygen because anaerobic methane oxidation is performed in nature by sulphate reducers^{22,23}:



and sulphate ions are likely to have been available from seawater or hydrothermal brines.

Methane-producing bacteria are ubiquitous in nature and must be regarded as one of the earliest species of organized life on Earth. They may, therefore also have played an important role in Precambrian times. The above model, summarized in Fig. 2, attempts to explain the as yet not understood variability of $^{13}\text{C}/^{12}\text{C}$ ratios in Precambrian reduced carbon. The extent of ^{13}C depletion would depend on the mixing ratio of normal CO_2 and CO_2 oxidized from CH_4 . An addition of 20% oxidized CO_2 to a CO_2 pool with an initial $\delta^{13}\text{C}$ of -7% would change the organic carbon $\delta^{13}\text{C}$ by 14% in favour of the light isotope, that is, the isotopic composition of -43% found in one graphite sample would have required an addition of $\sim 25\%$ of CO_2 which has been produced from CH_4 oxidation. Whether or not CH_4 and the CO_2 oxidized from it enters the CO_2 pool in which organic matter is metabolized depends on the specific situation of each basin. In shallow waters without stratification methane escapes without being oxidized; in well stratified waters methane is oxidized during diffusional exchange at density interfaces or alternatively at the sediment-water interface²⁰. Note that a relatively large fraction of our graphite samples (short non-iron formation horizons within the volcanic pile) is depleted in ^{13}C (three out of nine). In the graphites from large basins, no ^{13}C -depleted samples were found⁸. This may be due to differences in basin dimensions: in small basins stratified waters are more stable than in large basins.

The above processes are not necessarily confined to the small basins of the Archean of the Superior Province. They could have operated in other Precambrian areas where anomalous ^{13}C -depleted reduced carbon is found. The bulk of reduced Precambrian carbon has the same ^{13}C concentration as Phanerozoic kerogens and the ^{13}C -depleted reduced carbon might be explained by local effects. This reasoning argues against the interpretation of low ^{13}C concentrations in Precambrian organic matter as being indicative of special conditions in the Precambrian as compared with Phanerozoic times. The completely different environmental conditions (low O_2 partial pressure) obviously had not affected the basic processes of CO_2 assimilation.

We thank Teck Explorations Ltd for permission to publish the data.

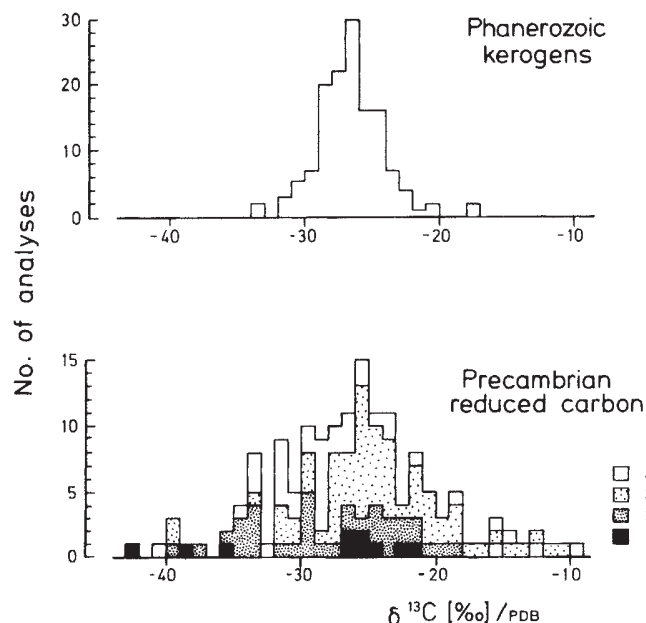


Fig. 1 Frequency distribution of $\delta^{13}\text{C}$ values of Precambrian reduced carbon compared with Phanerozoic kerogens. 1, This paper; 2, graphites from various Precambrian iron-formations^{4,5,8,11,12,14,15}; 3, disseminated carbon and carbon stringers from the middle Precambrian Transvaal dolomite¹⁶; 4, various Precambrian formations¹³. Phanerozoic kerogen values are unpublished data of the BGR stable isotope laboratory.

Received 1 September 1980; accepted 26 January 1981.

- Hoering, T. C. in *Researches in Geochemistry* (ed. Abelson, P. H.) (Wiley, New York, 1967).
- Rumble, D., Hoering, T. C. & Grew, E. S. *Yb. Carnegie Instn Wash.* **76**, 623 (1977).
- McKirdy, D. M. & Powell, T. G. *Geology* **2**, 591 (1974).
- Barghoorn, E. S., Knoll, A. H., Dembicki, H. & Meinschein, W. G. *Geochim. cosmochim. Acta* **41**, 425 (1977).
- Smith, J. W., Schopf, J. W. & Kaplan, I. R. *Geochim. cosmochim. Acta* **34**, 659 (1970).
- Degens, E. T. in *Organic Geochemistry* (eds Eglinton, G. & Murphy, M. T. J.) 304 (1969).
- Galimov, E. in *Kerogen Insoluble Organic Matter from Sedimentary Rocks* (ed. Durand, B.) 271 (Edition Technique, Paris, 1980).
- Goodwin, A. M., Monster, J. & Thode, H. G. *Econ. Geol.* **71**, 870 (1976).
- Krogh, T. E. & Davis, G. L. *Yb. Carnegie Instn Wash.* **70**, 241 (1971).
- Nunes, P. D. & Thurston, P. D. *Can. J. Earth Sci.* **17**, 710 (1980).
- Hoering, T. C. *Yb. Carnegie Instn Wash.* **61**, 190 (1962).
- Cloud, P. E. Jr, Gruner, J. W. & Hagen, H. *Science* **148**, 1713 (1965).
- Oehler, D. Z., Schopf, J. W. & Kvenvolden, K. A. *Science* **175**, 1246 (1972).
- Perry, E. C. Jr, Tan, F. C. & Morey, G. B. *Econ. Geol.* **68**, 1110 (1973).
- Schidlowski, M., Appel, P. W. U., Eichmann, R. & Junge, C. E. *Geochim. cosmochim. Acta* **43**, 189 (1979).
- Eichmann, R. & Schidlowski, M. *Geochim. cosmochim. Acta* **39**, 585 (1975).
- Schoell, M. & Hartmann, M. *Mar. Geol.* **14**, 1 (1973).
- Tietze, K., Geyh, M., Müller, H., Schröder, L. & Stahl, W. *Geol. Rdsch.* **69**, 452-472 (1980).
- Brooks, T. M., Bright, T. J., Bernard, B. B. & Schwab, C. R. *Limnol. Oceanogr.* **24**, 735 (1979).
- Sackett, W. M. *et al. Earth planet. Sci. Lett.* **44**, 73 (1979).
- Coleman, D. D., Risatti, J. & Schoell, M. (in preparation).
- Barnes, R. O. & Goldberg, E. D. *Geology* **4**, 297 (1976).
- Martens, C. S. & Berner, R. A. *Limnol. Oceanogr.* **22**, 10 (1977).
- Fallon, R. D., Havrils, S., Hanson, R. S. & Brock, T. D. *Limnol. Oceanogr.* **25**, 357 (1980).

Population density and body size in mammals

John Damuth

Committee on Evolutionary Biology, University of Chicago,
1103 E. 57 Street, Chicago, Illinois 60637, USA

There seems to be an inverse relationship between the size of an animal species and its local abundance. Here I describe the interspecific scaling of population density and body mass among mammalian primary consumers (herbivores, broadly defined). Density is related approximately reciprocally to individual metabolic requirements, indicating that the energy used by the local population of a species in the community is independent of its body size. I suggest that this is a more general rule of community structure.

Figure 1 shows the logarithm of mean population density plotted against the logarithm of mean adult body mass for 307 species of mammalian primary consumers. As far as possible, the density values represent 'ecological' densities, that is, those which apply to the habitat area actually used by the species. The

relationship is linear, with a slope of -0.75 , and species densities seem to be restricted to varying within about one order of magnitude from the value predicted by the regression line.

This analysis combines data from a wide variety of habitats throughout the world, and it is important that we know whether this overall pattern accurately reflects that found within individual communities; few, if any, mammal communities have been completely studied. To obtain a representative sample, I extracted from my data those sets yielding densities of three or more species within the same habitat type; these 'constructed' communities are shown in Table 1, with regression statistics. They include representatives of almost the whole range of habitat structures and primary productivity levels encountered by terrestrial mammals. None of the individual slopes, which vary from -0.56 to -0.95 , differ significantly from -0.75 , and Levene's test¹ and an analysis of covariance² reveal no significant differences in variance about the regression, in slope, or in intercept. Pooling the data from the communities gives an estimated slope of -0.70 , which is not significantly different from the value obtained for the overall regression (t -test). Thus, the overall trend gives a reasonable representation of that which we are likely to find within individual communities.

Knowledge of population density scaling in communities allows us to consider the relative energy use of species among primary consumers of different sizes. Individual basal metabolic requirements are related to body mass by the power of 0.75 (refs 3, 4). Estimates of the metabolic requirements of free-living mammals in natural habitats roughly parallel basal requirements, but at a higher level, varying between ~ 1.5 and 3.0 times basal values⁵⁻¹³. Thus they will also be related to body mass by the power of ~ 0.75 . The energy used by the local population of a species equals the population density (D) multiplied by individual metabolic requirements (R), which yields the following relationship to body mass (W): $DR \propto W^{-0.75} W^{0.75}$. The exponents of W cancel each other, which gives the important result that the amount of energy that a species population uses in the community is independent of its body size. No mammal herbivore species, on an ecological time scale, has an energetic advantage over any other solely as a result of size differences.

There are two important corollaries of this relationship. First, because secondary productivity for a given amount of assimilated energy is independent of body size in mammals (Kleiber's Law)³, it follows from the above that the secondary productivity of a herbivore species' local population, and hence the energy that it yields to the next higher trophic level, is also independent of body size. Second, the standing-crop biomass of a species (population density multiplied by body mass) is positively related to body mass by the ~ 0.25 power; this is a

Table 1 Habitat types and constructed communities

Community/habitat type	a	s.e.	r	b	n
Sonoran desert, USA	-0.63	0.16	-0.76	3.37	14
Mesquite grassland, USA and Mexico	-0.56	0.086	-0.95	3.73	7
Boreal and subalpine forest, N. America	-0.79	0.080	-0.95	4.43	12
Lowland tropical rainforest, Malaysia	-0.60	0.089	-0.90	2.61	13
Transvaal lowveld (woodland-savanna), S. Africa	-0.61	0.089	-0.93	3.78	10
Mixed temperate forest, Poland	-0.79	0.080	-0.97	4.33	8
Temperate grassland, USA	-0.67	0.13	-0.92	3.28	7
Tropical grassland, Rwenzori N.P., Uganda	-0.79	0.17	-0.81	3.59	13
Tropical grassland, Sri Lanka*	-0.79	0.19	-0.92	4.59	5
Ichu grassland, Altiplano, Peru*	-0.82	0.34	-0.99	3.93	4
High arctic tundra, Canada*	-0.95	0.21	-0.96	4.71	4
Sub-arctic birch forest and meadows, Norway*	-0.83	0.72	-0.99	4.41	3
Southern pine-hardwood forest, USA*	-0.91	0.46	-0.89	4.23	3
Northern hardwoods, USA*	-0.57	0.11	-0.98	3.11	3
Oaks and chapparal, USA*	-0.72	0.021	-0.99	4.83	3
Means of statistics	-0.74	0.20	-0.93	3.93	—
Pooled data from above regressions	-0.70	0.040	-0.86	3.71	109

Statistics are for standard least-squares regression equations: $\log D = a(\log W) + b$; s.e., standard error of the slope (a); r, correlation coefficient; n, number of species. Note that in the pooled regression, as some species are found in more than one habitat type, $n = 109$ includes only 92 separate species.

* Due to small sample size and/or a single point at one end of the size range, the individual values for these regressions are not very reliable.

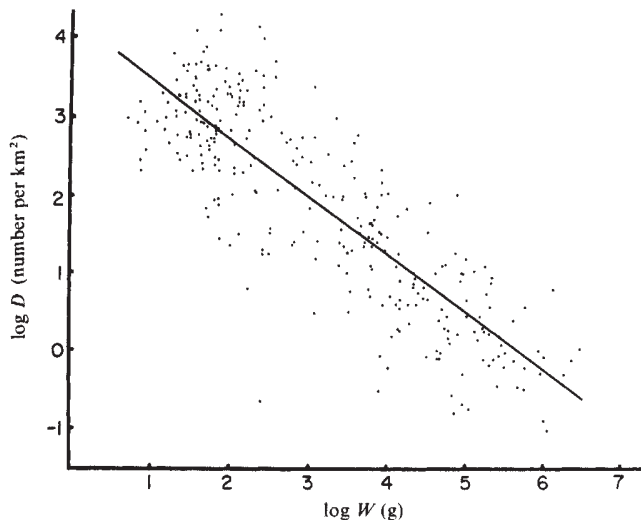


Fig. 1 Population density (D) compared with the mean adult body mass (W) for 307 mammal primary consumers; each point represents one species. Density values for each species are the mean of the means from each locality from which data were reported for the particular species. (Data are from the literature for the years 1950–79, derived from 115 journals and numerous books, ~650 references in all.) The line represents the least-squares regression line, $\log D = -0.75 (\log W) + 4.23$; $r = -0.86$, standard error of the slope = 0.026.

quantitative measure of a qualitative relationship that has been known for some time but only sporadically appreciated¹⁴. Species of small mammals are able to produce, on average, the same amount of biomass over time as do species of large mammals, whereas at a given moment their standing-crop biomass is considerably less, because the population turnover rates and individual growth rates per unit weight of small species are much greater^{15–17}.

The widespread occurrence of an approximately reciprocal relationship between population density and individual metabolic requirements among mammalian herbivores suggests that a general principle is involved. The independence of species energy control and body size revealed by this reciprocal relationship implies that random environmental fluctuations and interspecific competition act over evolutionary time to keep energy control of all species within similar bounds. It is unlikely that the occurrence of competition among species of different body sizes is restricted to herbivorous mammals. Along with the ubiquity of allometric scaling of metabolic rate with body size⁴, this suggests that a reciprocal relationship, and particularly a value of ~ -0.75 , characterizes a broader range of taxa and trophic levels. These points will be discussed more fully elsewhere.

I thank I. L. Heisler, V. C. Maiorana, D. M. Raup, V. L. Roth and L. Van Valen for comments, and acknowledge support from the University of Chicago and Field Museum of Natural History.

Received 6 October 1980; accepted 27 January 1981.

1. Van Valen, L. *Evol. Theory* **4**, 33–44 (1978).
2. Brownlee, K. A. *Statistical Theory and Methodology in Science and Engineering* 2nd edn (Wiley, New York, 1965).
3. Kleiber, M. *The Fire of Life* 2nd edn (Krieger, New York, 1975).
4. Hemmingen, A. M. *Rep. Steno meml Hosp.* **4**, 7–58 (1950).
5. Brody, S. *Bioenergetics and Growth* (Reinhold, New York, 1945).
6. Chew, R. M. & Chew, A. E. *Ecol. Monogr.* **40**, 1–21 (1970).
7. Gessaman, J. A. in *Ecological Energetics of Homeotherms* (ed. Gessaman, J. A.) (Utah University Press, 1973).
8. King, J. R. in *Avian Energetics* (ed. Paynter, R. A. Jr) (Nuttall Ornithological Club, Massachusetts, 1974).
9. McKay, G. M. *Smithson. Contr. Zool.* **125**, 1–113 (1973).
10. Moen, A. N. *Wildlife Ecology: an Analytical Approach* (Freeman, San Francisco, 1973).
11. Mullen, R. K. *Comp. Biochem. Physiol.* **39**, 379–390 (1971).
12. Mullen, R. K. & Chew, R. M. *Ecology* **54**, 633–637 (1973).
13. Nagy, K. A. & Milton, K. *Ecology* **60**, 475–480 (1979).
14. Odum, E. P. *Fundamentals of Ecology* 3rd edn (Saunders, Philadelphia, 1971).
15. Fenchel, T. *Oecologia* **14**, 317–326 (1974).
16. Western, D. *Afr. J. Ecol.* **17**, 185–204 (1978).
17. Case, T. J. Q. *Rev. Biol.* **53**, 243–282 (1978).

Competitive ability influences habitat choice in marine invertebrates

Richard K. Grosberg

Department of Biology, Yale University, 260 Whitney Avenue, PO Box 6666, New Haven, Connecticut 06511, USA

Patterns of distribution and abundance of sessile marine epibenthic invertebrates are controlled by three factors: (1) the presence and abundance of larvae which are competent to settle, (2) the choice of settling sites by recruiting larvae, and (3) the biotic and physical events occurring during and after settlement. Although there is much information on the distribution of larvae and seasons of recruitment^{1–3}, substratum selection^{4–6} and post-settlement events^{7–15}, very little is known of the ecological and evolutionary relationships between these factors^{5,16–18}. Natural selection acts on entire life cycles, thus information about these relationships is essential for understanding patterns of recruitment and survival. For example, sessile organisms can modify the course of post-recruitment events by selective settlement and directional growth¹⁹. Here, I present evidence that the larvae of several taxa of marine invertebrates avoid substrata where there is a high probability of death caused by a superior spatial competitor.

I found, as did Grave²⁰, that in the Eel Pond at Woods Hole, Massachusetts (USA), the two compound ascidians *Botryllus schlosseri* (Pallas) Savigny and *Botrylloides leachii* (Savigny) cover, by area, over 50%, and sometimes 100%, of hard substrata during the spring and summer. This period includes the season of recruitment for most other colonizers. Thus, settling larvae are likely to contact a botryllid ascidian at some time during their lives. A common result of contacts between colonists is overgrowth of one organism by another^{21–29}. In this way, overgrowth is a consequence of competition for space (and perhaps other resources³⁰), and therefore can be an important factor in the distribution and abundance of encrusting species.

B. schlosseri is the most successful overgrower (along with *B. leachii*) in the Eel Pond (Table 1) and is rarely overgrown by other species. As it is such an important member of the Eel Pond epibenthic assemblage (considering the post-settlement events of overgrowth and coverage), I investigated the effects of its density on the recruitment of other taxa.

B. schlosseri collected from the Eel Pond were placed in finger-bowls where they released larvae which were then transferred to 10-ml polycarbonate dishes. Two hours after settlement on the sides of the dishes, the juveniles were teased off and re-attached to 5 × 5 cm glass plates. I set up three densities of juvenile *Botryllus* with three replicates of each: (1) 15 regularly spaced *Botryllus* juveniles, with all subsequent botryllid settlement removed daily; (2) 5 regularly spaced *Botryllus* juveniles with all subsequent botryllid recruits removed daily; and (3) no *Botryllus* juveniles with daily removal of all botryllid recruits. Three additional glass plates which initially carried no juveniles, but on which all recruits were allowed to remain, served as controls.

Experimental and control plates were positioned horizontally in a Latin square arrangement³¹ below a floating dock at 1 m depth. The experiment was started on 20 July, 24 h after the larvae had been transplanted. The lower side of each plate (which carried the juvenile *Botryllus*) was censused non-destructively until 28 July on a daily basis—this allowed me to distinguish between failure to settle and all but the earliest post-settlement mortality.

Figure 1 shows the numbers of each species that settled on the glass plates. To determine if there were any differences between each of the three density treatments and the controls, the data for each species were analysed using the null hypothesis that

settlement was uniform in each case. χ^2 tests unambiguously placed the species in two groups: (1) selective species or those for which settlement was not uniform between treatments (Fig. 1a, all $P < 0.005$) and (2) non-selective species or those for which there were no significant differences in settlement between density treatments (all $P > 0.2$). When settlement on the plates carrying five *B. schlosseri* was compared with that on plates where all botryllids had been removed, settlement was uniform for all species ($P > 0.5$). On control plates, where cumulative *B. schlosseri* settlement reached 15 per plate on the third day of the experiment, settlement of selective species was intermediate between that on plates carrying 15 *B. schlosseri* and those carrying none or five juveniles. This implies that the threshold for detection of *B. schlosseri* by selective species lies somewhere between 5 and 15 resident *Botryllus* juveniles.

There is a clear relationship between patterns of recruitment and a species' susceptibility to overgrowth. The species which did not settle on plates where *Botryllus* number was 15 (selective

species) were also overgrown by *Botryllus* in 1,008 of 1,048 contacts (96.2%). These species are the first nine of Table 1. In contrast, those species which settled uniformly, regardless of *B. schlosseri* density (non-selective species), were rarely overgrown by *Botryllus* (171 of 961 contacts or 17.1%). These species include the remaining 10 of Table 1.

The mechanisms by which settlement is decreased for some taxa, but not for others, are unclear. Allelopathy^{22,32,33} cannot account for these results, for when selective species do settle on the plates (for example, at low *Botryllus* densities), they often settle immediately adjacent to the competitive dominant. In addition, because *Botryllus* cannot ingest particles as large as most recruiting larvae^{34,35}, selective predation does not explain the results. It is more likely that the settling larvae of some taxa can recognise the density of resident *Botryllus*, and reject substrata when that density exceeds a certain threshold.

Selective settlement has been observed in a number of taxa, but it is not understood which factors make particular substrata

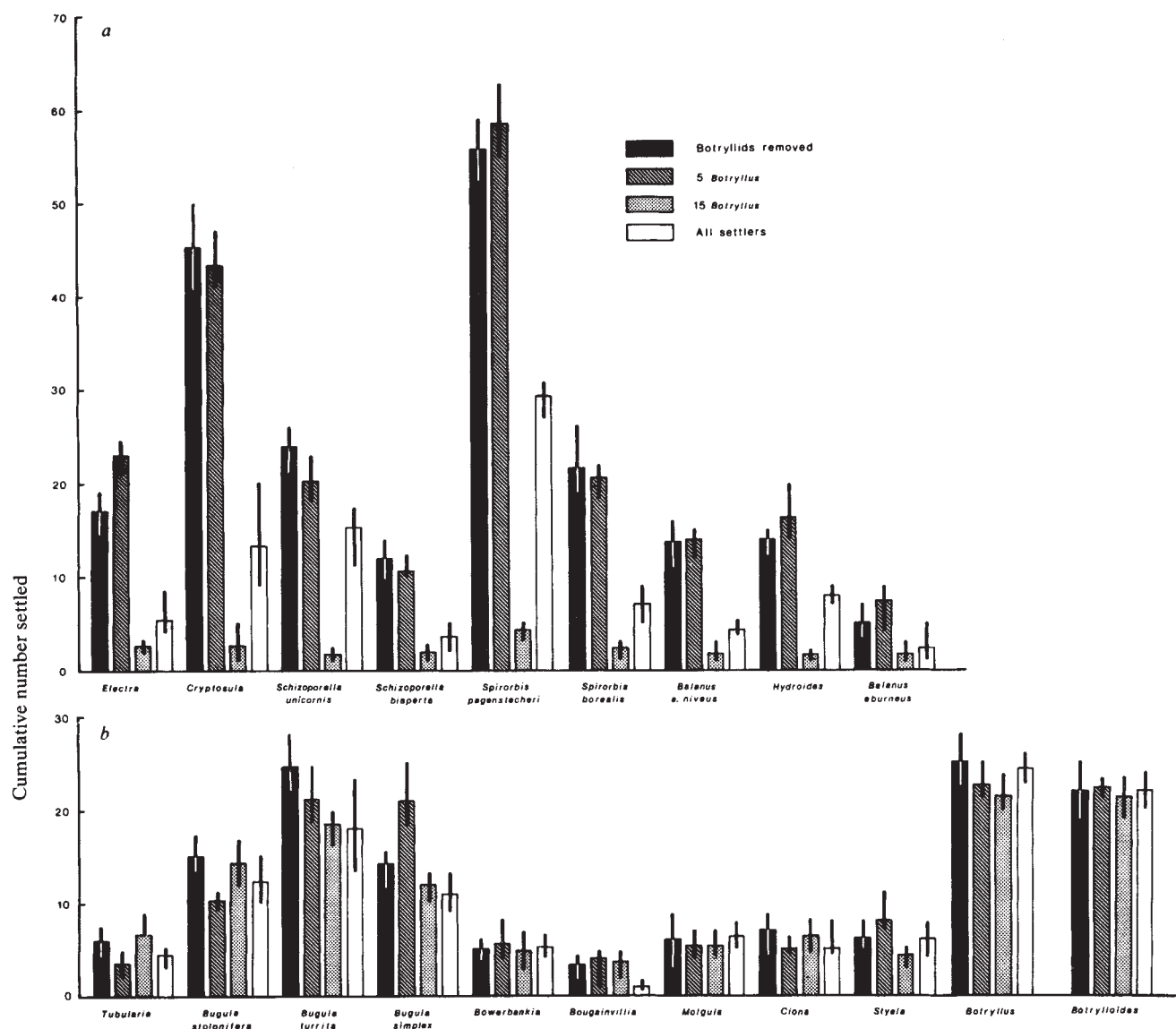


Fig. 1 The cumulative settlement for all species which settled during the experiment. Bars represent the mean settlement for the three replicates of each density treatment and the controls. Lines indicate the range of settlement values. The plates were arranged in a Latin square so that plate position effects could be detected: the χ^2 analyses indicated that there were no significant differences between replicates, thus there were no plate position effects. Of the 60 transplanted larvae, only one died during the experiment. Any settler not immediately identified was allowed to grow after the term of the experiment until identification was possible. At the end of the experiment, none of the colonists covered $> 2 \text{ mm}^2$, thus coverage was low. *a*, Data for species which showed a settlement effect at different *B. schlosseri* densities. These are termed selective species—this group includes four encrusting bryozoans, three tube-building polychaetes and two barnacles. All these species have feeding structures close to the substratum, thus easily obstructed by overgrowth. *b*, Data for those species which showed no significant differences in settlement at different *Botryllus* densities. These are termed non-selective species—this group includes four arborescent bryozoans, two stoloniferous hydroids, three solitary ascidians and two colonial botryllid ascidians. With the exception of the botryllids, all members of this group have elevated feeding structures and therefore are generally resistant to the effects of overgrowth (see refs 23, 28).

Table 1 Results of contacts between *Botryllus schlosseri* and the adults of the species which recruited during the period 20–28 July 1979

Overgrown	Wins	Ties	Total	Wins/total (%)	Losses/total (%)
<i>Electra crustulenta</i>	76	0	78	76/78 (97.4%)	2/78 (2.6%)
<i>Cryptosula pallasiana</i>	159	0	159	159/159 (100%)	0/159 (0%)
<i>Schizoporella biaperta</i>	43	0	43	43/43 (100%)	0/43 (0%)
<i>Schizoporella unicornis</i>	58	0	63	58/63 (92.1%)	5/63 (7.9%)
<i>Spirorbis pagenstecheri</i>	382	0	382	382/382 (100%)	0/382 (0%)
<i>Spirorbis borealis</i>	118	13	131	118/131 (90.1%)	0/131 (0%)
<i>Balanus amphitrite niveus</i>	79	10	89	79/89 (88.8%)	0/89 (0%)
<i>Balanus eburneus</i>	47	5	52	47/52 (90.4%)	0/52 (0%)
<i>Hydroides dianthus</i>	46	5	51	46/51 (90.2%)	0/51 (0%)
<i>Botrylloides diegensis</i>	147	108	391	147/391 (37.6%)	136/391 (34.8%)
<i>Tubularia larynx</i>	3	87	84	3/87 (3.5%)	0/87 (0%)
<i>Bougainvillia</i> sp.	2	32	34	2/34 (5.9%)	0/34 (0%)
<i>Bugula stolonifera</i>	2	59	61	2/61 (3.3%)	0/61 (0%)
<i>Bugula turrita</i>	4	76	80	4/80 (5.0%)	0/80 (0%)
<i>Bugula simplex</i>	3	99	102	3/102 (2.9%)	0/102 (0%)
<i>Bowerbankia gracilis</i>	6	64	70	6/70 (8.6%)	0/70 (0%)
<i>Molgula manhattensis</i>	1	26	27	1/27 (3.7%)	0/27 (0%)
<i>Ciona intestinalis</i>	1	15	16	1/16 (6.3%)	0/16 (0%)
<i>Stryla clava</i>	2	95	97	2/97 (2.1%)	0/97 (0%)

These data were collected during the summer of 1979 from contacts between animals living on algae, floating docks, rocks and glass plates in the Eel Pond. All contacts were followed from their initiation until their resolution by overgrowth or mutual retreat. Total contacts equal wins + losses + ties. A win is scored when *Botryllus* overtops and kills >50% of a colonial organism, or entirely kills a solitary organism. A tie is scored when growth of both participants ceases along a contact, or when *Botryllus* simply grows around another colonizer without perceptible harm to either. Although *Botryllus* infrequently overgrows certain taxa, it is rarely overgrown itself by any taxa. Nevertheless, the canopies of arborescent bryozoans and hydroids sometimes shade underlying compound ascidians.

(or habitats) better than others. Notable exceptions include: (1) the gregarious settlement of some barnacles which apparently ensures cross-fertilization and viability of eggs^{6,33,36}; (2) the preferential settlement of predators, commensals and parasites near their prey, partners and hosts, respectively⁵; (3) the selection by some bryozoans and polychaetes of particular algae (or positions on these algae) which correlates with patterns of survival of juveniles and adults^{17,18,37}; and (4) the gregarious settlement of some arborescent bryozoans which enhances their interspecific competitive ability³⁸.

The results from the Eel Pond indicate that some larvae can distinguish substrata where post-settlement events are likely to kill them. Apparently, these larvae avoid encounters with the competitive dominant, *B. schlosseri*, by settling away from dense stands of it. In comparison, the larvae that do settle where *B. schlosseri* is abundant are those most immune to overgrowth by it. Taken together, the data illustrate that competition during the post-settlement phase of an organism's life cycle can strongly influence the evolution of habitat selection.

This research was supported by the Lerner Fund for Marine Research of the American Museum of Natural History. The Woods Hole Oceanographic Institution provided laboratory space. I thank L. Buss, R. Harrison, R. Hartigan, B. Keller, B. Kennedy and N. Knowlton for their helpful criticisms.

Received 29 October 1980; accepted 28 January 1981.

- MacDougall, K. D. *Ecol. Monogr.* **13**, 321–374 (1943).
- Thorson, G. *Neth. J. Sea Res.* **3**, 267–293 (1966).
- Osman, R. W. *Ecol. Monogr.* **47**, 37–63 (1977).
- Meadows, P. S. & Campbell, J. I. *Adv. mar. Biol.* **10**, 271–382 (1972).
- Scheltzema, R. S. *Thalassia jugosl.* **10**, 263–296 (1974).
- Crisp, D. J. in *Adaptations to Environment: Essays on the Physiology of Marine Organisms* (ed. Newell, R. C.) 83–124 (Butterworths, London, 1976).
- Connell, J. H. *Ecology* **42**, 710–723 (1961).
- Connell, J. H. *Ecol. Monogr.* **31**, 61–104 (1961).
- Paine, R. T. *Ecology* **50**, 950–961 (1969).
- Paine, R. T. *Oecologia* **15**, 93–120 (1974).
- Dayton, P. K. *Ecol. Monogr.* **41**, 315–389 (1971).
- Connell, J. H. & Slatyer, R. O. *Am. Nat.* **111**, 1119–1144 (1977).
- Menge, B. A. *Ecol. Monogr.* **46**, 355–393 (1976).
- Sutherland, J. P. & Karlson, R. H. *Ecol. Monogr.* **47**, 425–446 (1977).
- Buss, L. W. & Jackson, B. C. *Am. Nat.* **113**, 223–234 (1979).
- Woodin, S. A. *J. mar. Res.* **34**, 25–41 (1976).
- Mackay, T. F. C. & Doyle, R. W. *Heredity* **40**, 1–12 (1978).
- Hayward, P. J. & Harvey, P. H. *J. mar. biol. Ass. U.K.* **54**, 677–684 (1974).
- Buss, L. W. in *Biology and Systematics of Colonial Organisms* (ed. Larwood, G. P.) 459–497 (Academic, London, 1979).
- Grave, B. H. *Biol. Bull.* **65**, 375–386 (1933).
- Gordon, D. P. *J. nat. Hist.* **6**, 503–514 (1972).
- Bryan, P. G. *Micronesica* **9**, 237–242 (1973).
- Stebbing, A. R. D. *J. mar. biol. Ass. U.K.* **53**, 247–261 (1973).
- Stebbing, A. R. D. in *Living and Fossil Bryozoa* (ed. Larwood, G. P.) 173–183 (Academic, London, 1973).

- Jackson, J. B. C. *Am. Nat.* **111**, 743–767 (1977).
- Jackson, J. B. C. *J. Anim. Ecol.* **48**, 805–823 (1979).
- Karlson, R. H. *J. exp. mar. Biol. Ecol.* **31**, 225–239 (1978).
- Russ, G. R. *J. exp. mar. Biol. Ecol.* **42**, 55–69 (1980).
- Sebens, K. P. *Ecology* (in the press).
- Buss, L. W. *Nature* **281**, 475–477 (1980).
- Sokal, R. R. & Rohlf, F. J. *Biometry* (Freeman, San Francisco, 1969).
- Goodbody, I. *Nature* **190**, 282–283 (1961).
- Knight-Jones, E. W. & Moyle, J. *Symp. Soc. exp. Biol.* **15**, 72–95 (1961).
- Lorenzoni, G. G. *Atti Ist. veneto* **126**, 147–168 (1968).
- Milkman, R. *Biol. Bull.* **132**, 229–243 (1967).
- Barnes, H. & Crisp, D. J. *J. mar. biol. Ass. U.K.* **35**, 631–639 (1956).
- Knight-Jones, E. W., Knight-Jones, P. & Al-Olgily, S. M. in *9th Eur. mar. biol. Symp.* (ed. Barnes, H.) 539–561 (Aberdeen University Press, 1975).
- Buss, L. (in review).

Fertilization potential in golden hamster eggs consists of recurring hyperpolarizations

Shun-ichi Miyazaki & Yukio Igusa

Department of Physiology, Jichi Medical School, Tochigi-ken 329-04, Japan

The fertilization potential, or activation potential, has been demonstrated in the eggs of various species¹, and it has been shown to block polyspermy in echinoderm^{2,3}, echinuran⁴ and frog⁵ eggs, but no studies have been reported of electrical phenomena occurring when mammalian eggs are fertilized. We report here the fertilization potential of golden hamster eggs *in vitro*. To correlate the change of potential with the interaction between sperm and egg, only one sperm was attached to each egg. We found that a sperm induces recurring hyperpolarizations, constituting a fertilization potential which differs from that in the eggs of other species.

Mature eggs were collected from the oviducts of superovulated females⁶ and freed from the surrounding cumulus cells and zona pellucida by sequential treatment with 0.1% hyaluronidase (~5 min at 23–26 °C) and 0.1% trypsin (4–5 min) in medium used for rat eggs⁷. Removal of the zona allows sperm to reach the plasma membrane soon after insemination. Spermatozoa obtained from the cauda epididymis were activated with acrosome reaction by incubation in modified Tyrode solution containing 20% human serum⁸ and 1% bovine adrenal gland extract⁹ at 37 °C for 4 h. All treatments and experiments were performed in a 0.4 ml drop of medium placed in a plastic Petri dish and covered with paraffin oil. A glass microelectrode was

inserted in an egg viewed in a phase-contrast inverted microscope and a small amount of the sperm suspension was added to the egg medium at 30–32 °C (final concentration 2,000–5,000 sperm per 0.4 ml). Electrode resistances were in the range 40 to 80 M Ω , and leakage current of a preamplifier was <2 pA. Each experimental egg was transferred quickly to sperm-free medium after withdrawing the electrode, and was incubated for 4 h to determine sperm entry by observation of male pronuclei.

The resting potential of unfertilized eggs averaged -29 ± 7 mV (mean $\pm$ s.d., $n = 120$), but 7% of the values were between -40 and -50 mV. The current–voltage relationship (I – V curve) was linear for potentials of between -100 and 0 mV, and the effective membrane resistance averaged 142 ± 65 M Ω ($n = 98$). The more negative resting potential tended to be associated with the larger membrane resistance, suggesting that there was a substantial leakage due to electrode impalement in most cases. The largest resting potential was -50 mV with a membrane resistance of 400 M Ω , which is estimated to be 66 k Ω cm 2 , assuming that the egg is a sphere of 72 μ m diameter.

Figure 1A shows the fertilization potential in the case of the severe polyspermy which usually occurs in zona-free eggs. The major potential change is a repetitive, transient hyperpolarization (H-response, HR) associated with an increase in membrane conductance. There was no observed HR during a 10-min recording before insemination. The first HR had the following characteristics: peak potential, -72 ± 8 mV (mean $\pm$ s.d., $n = 53$); duration, 14 ± 3 s; membrane resistance near the peak, 28 ± 9 M Ω . The membrane conductance became 4.5 ± 1.8 times larger than that just before insemination. Successive HRs differed in minor ways: they became slightly longer and the peak became slightly less negative with a smaller increase in conductance (Fig. 1A). Summation of HRs was never observed, but they appeared in an all-or-none manner. The repetitive HRs persisted at almost constant intervals for at least 15 min while they were recorded, during which time some additional sperm attached to the egg after the initial rush of sperm. The resting membrane potential shifted gradually in the hyperpolarizing direction to ~ -40 mV, and the conductance increased while the I – V curve remained linear. A small depolarization which appeared immediately after insemination was not induced by sperm, but was caused by serum and/or adrenal gland extract in the sperm incubation medium (Fig. 1Ba, b).

In the conditions used in our experiments, fertilization did not occur in 30% of impaled eggs. Table 1 indicates that the occurrence of HR(s) was associated with sperm entry into an egg

Table 1 Correlation between HR(s) and fertilization

	HR (+)*	HR (–)
Sperm entry (+)*	44 eggs	4†
Sperm entry (–)	3‡	20

Sperm entry was determined by observation of male pronucleus or decondensation of sperm head.

* Occurrence of HR or sperm entry, irrespective of the number.

† Additional sperm attached to the egg after the electrode was removed. It is likely that these sperm entered the egg.

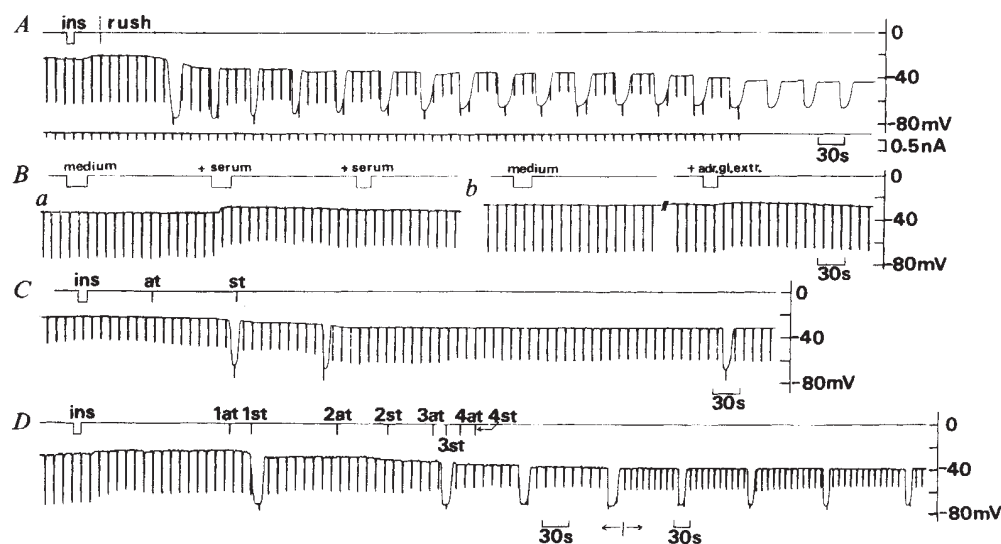
‡ The eggs had been damaged, thus decondensation of the sperm head might be prevented.

and that, conversely, sperm failed to enter eggs that had not shown any HR. When sperm with insufficient pretreatment were used, no HR was induced; these sperm have intact acrosomes and stick to the surface but never fuse with the egg¹⁰.

To correlate the occurrence of HR(s) with the interaction between sperm and egg, the number of sperm reaching an egg was restricted by reducing the concentration (1,000–2,000 sperm per 0.4 ml) and by placing a glass half-ring around the egg as a barrier. The first sperm reached and attached to the egg several minutes after insemination ('at' in Fig. 1). Then, using a micromanipulator, we added another half-ring to form a closed ring. Active flagellar motion decreased fairly suddenly, 10–90 s after the sperm had attached. The sperm then stopped moving completely in 10–15 s, becoming straight ('st' in Fig. 1). Table 2 shows that the first HR began in most cases within 10 s before or after the stop of the first sperm: there was a fairly good correlation between the two events.

In Fig. 1C, only one sperm bound to the egg throughout the experiment (Fig. 2A), and it was incorporated into the egg to form a male pronucleus (Fig. 2B). Three HRs appeared during 11 min after the sperm had stopped moving. In 10 out of 13 eggs penetrated by a single sperm, 2–4 HRs appeared within 2 min after the sperm had stopped. The next HR tended to occur after a pause of several minutes, as in Fig. 1C. Another three eggs behaved differently: one showed a continuous series of small HRs and two showed only one. The recurring HRs were also recorded in a zona-intact egg. In the case of Fig. 1D, four sperm attached and stopped separately from each other, and all were found to be incorporated into the egg. The stops of flagellar motion of additional sperm did not always coincide with the occurrence of a HR. However, multiple sperm entries caused a

Fig. 1 Fertilization potential induced by 16(A), one (C) and four sperm (D). To monitor a change in the membrane conductance, repetitive current pulses were applied through the intracellular electrode by means of a bridge circuit, with a current strength of 0.3 nA in A (bottom trace), and 0.2, 0.5, 0.5 and 0.15 nA in Ba to D, respectively. The upper trace shows events which were marked electrically by the observer, and also indicates zero potential: ins, insemination; rush, rush of sperm to attach to the egg surface; 1–4 at and st, attachment and stop of the first to fourth sperm. In B, a small depolarization associated with conductance increase was induced by addition of 20 μ l of sperm incubation medium containing 20% human serum (Ba) or 1% bovine adrenal gland extract (Bb). The incubation medium itself had no effect. The egg was insensitive to the second application of serum (Ba). There was a 3-min interruption in the middle of Bb. Concentrations of major ions in the egg medium were as follows (mM): Na $^+$, 119; K $^+$, 5.9; Ca $^{2+}$, 1.7; Mg $^{2+}$, 1.2; Cl $^-$, 102; HCO $_3^-$, 26 (equilibrated with CO $_2$ at pH 7.4).



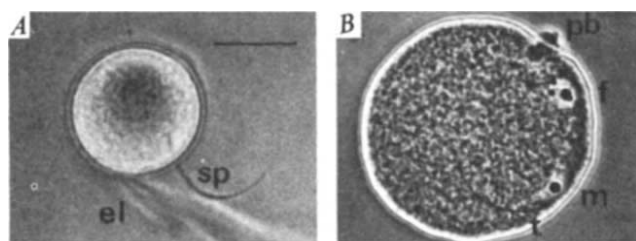


Fig. 2 The egg from which the record of Fig. 1C was obtained. **A**, After recording of potentials; **B**, after incubation for 4 h. The egg was pressed with a cover slip in **B** to visualize pronuclei. Scale bar for **A** and **B**, 50 μ m. sp, Sperm; el, electrode; pb, second polar body; f, female pronucleus; m, male pronucleus; t, sperm tail.

continuous series of HRs without any long pause. The series of HRs persisted although no more sperm became attached.

We conclude that the fertilization potential in hamster eggs consists of recurring hyperpolarizations. In our experiments, the time of recording the potentials after insemination corresponded to the period from contact between sperm and egg to, at most, the incorporation of the entire sperm head into the egg cytoplasm^{10,11}. The HR first appeared at about the time of the first sperm stop, when sperm and egg are likely to fuse¹⁰. The periodic appearance of HRs thereafter may be a consequence of a pre-existing ability of the egg, because similarly shaped HRs appear at almost constant intervals even without correlation with each sperm-egg fusion, as shown in the case of polyspermy (Fig. 1A, D). Repetitive hyperpolarizing responses each with durations of several to 20 s have been found in fibroblastic L cells^{12,13}, macrophages^{14,15} and sympathetic ganglion cells¹⁶. These HRs are due to an increase in the K permeability^{12,15} caused by an increase in intracellular Ca^{2+} ($[\text{Ca}^{2+}]_i$)^{14,16,17}. The HRs in hamster eggs had a reversal potential between -85 and -80 mV (our unpublished results), suggesting an increase in K permeability. The HRs in the above cells occur either spontaneously or by electrical, chemical or mechanical stimulations, whereas those in hamster eggs were induced neither by injecting current up to 5 nA nor by pushing the egg with a glass capillary, but they were induced by sperm about to be incorporated into the egg. It has been shown in sea urchin eggs^{18,19} that one of the early events on fertilization is the release of Ca^{2+} from intracellular stores and the consequent increase in $[\text{Ca}^{2+}]_i$ mediates several changes such as the discharge of the cortical granules. The HR in hamster eggs may reflect a cyclic change in $[\text{Ca}^{2+}]_i$ during incorporation of the sperm into the egg, although the functional significance is unknown.

In echinoderm and echiuran eggs, the membrane potential shifts to positive values during the early phase of the fertilization potential, due to an increase in membrane permeability to Na^+ and possibly Ca^{2+} (ref. 1). In the medaka fish, *Oryzias latipes*, fertilization triggers a depolarization of a few mV followed by a longer hyperpolarizing phase with an amplitude of ~30 mV, when the resting potential is more negative than -20 mV (ref. 20). In hamster eggs we could not record any depolarization preceding the HR in recording conditions that were able to detect a change in potential of at least 1 mV. This was also the case when the membrane potential was preliminarily shifted to

-64 to -68 mV by applying current. In one of four such cases, the membrane potential and effective resistance were -65 mV and 310 M Ω , respectively. We have proposed an equivalent electrical circuit for the egg plasma membrane²¹ analogous to circuits for subsynaptic membranes of nerve cells²². If one assumes a possible 1-mV depolarization caused by a shunting conductance (g_s) with a reversal potential of -20 mV, the g_s in the above egg is calculated as 72 pS. If the potential is clamped at -65 mV, the g_s causes a membrane current of 3.2 pA. Thus, the depolarizing response, if any, is negligibly small. Furthermore, preliminary experiments indicated that the potential-dependent fertilization block shown in echinoderm^{2,3}, echiuran⁴ and frog⁵ eggs does not exist in the hamster egg. These differences stimulate further experimentation in mammalian eggs.

We thank Drs A. Hanada and Y. Hirao for teaching us the techniques of fertilization *in vitro* and for valuable advice, Dr S. Ozawa for criticism, and Miss Y. Hodota for technical assistance.

Received 8 December 1980; accepted 3 February 1981.

- Hagiwara, S. & Jaffe, L. A. *Rev. Biophys. Bioengng* **8**, 385-416 (1979).
- Jaffe, L. A. *Nature* **261**, 68-71 (1976).
- Miyazaki, S. & Hirai, S. *Dev. Biol.* **70**, 327-340 (1979).
- Gould-Somero, M., Jaffe, L. A. & Holland, L. J. *Cell Biol.* **82**, 426-440 (1979).
- Cross, N. L. & Elinson, R. P. *Dev. Biol.* **75**, 187-198 (1980).
- Yanagimachi, R. *J. Reprod. Fert.* **18**, 275-286 (1969).
- Toyoda, Y. & Chang, M. C. *J. Reprod. Fert.* **36**, 9-22 (1974).
- Yanagimachi, R. *Biol. Reprod.* **3**, 147-153 (1970).
- Bavister, B. D., Yanagimachi, R. & Teichman, R. J. *Biol. Reprod.* **14**, 219-221 (1976).
- Yanagimachi, R. in *Current Topics in Developmental Biology* Vol. 12 (ed. Monroy, A.) 83-105 (Academic, New York, 1978).
- Yanagimachi, R. & Noda, Y. D. *Am. J. Anat.* **128**, 429-462 (1970).
- Nelson, P. G., Peacock, J. & Minna, J. J. *gen. Physiol.* **60**, 58-71 (1972).
- Okada, Y., Doida, Y., Roy, G., Tsuchiya, W., Inoue, K. & Inoue, A. *J. Membrane Biol.* **35**, 319-335 (1977).
- Gallin, E. K., Wiederhold, M. L., Lipsky, P. E. & Rosenthal, A. S. *J. cell. Physiol.* **86**, 653-662 (1975).
- Dos Reis, G. A. & Oliveira-Castro, G. M. *Biochim. biophys. Acta* **469**, 257-263 (1977).
- Kuba, K. & Nishi, S. *J. Neurophysiol.* **39**, 547-563 (1976).
- Okada, Y., Tsuchiya, W. & Inoue, A. *J. Membrane Biol.* **47**, 357-376 (1979).
- Steinhardt, R. A. & Epel, D. *Proc. natn. Acad. Sci. U.S.A.* **71**, 1915-1919 (1974).
- Steinhardt, R., Zucker, R. & Schatten, G. *Dev. Biol.* **58**, 185-196 (1977).
- Nuccitelli, R. *Dev. Biol.* **76**, 483-498 (1980).
- Miyazaki, S. *Dev. Biol.* **70**, 341-354 (1979).
- Eccles, J. C. *The Physiology of Synapses* (Springer, New York, 1964).

Electrical coupling between fibre cells in amphibian and cephalopod lenses

T. J. C. Jacob & G. Duncan

School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK

The lenses of vertebrate and cephalopod eyes differ ontogenetically and in other respects. The vertebrate lens, derived from a single cell type, consists mainly of long fibre cells continuously produced by division and elongation of columnar epithelial cells near the lens equator. Almost 50% of the fibre cell surface consists of junctional complexes^{1,2} and the internal resistance, from point to point within the lens, is low compared with the surface membrane resistance³. Thus the vertebrate lens is expected to behave as a well coupled syncytial system⁴. The cephalopod lens, however, is formed by the fusion of two distinct cell types⁵; the anterior segment has the same ontogenetic origin as the cornea but the posterior segment shares a common origin with the retina, and the plane of contact of the two cell types can be seen in light-microscope sections⁵. Most of the lens is composed of long fibre cells similar in appearance to those found in the vertebrate lens, and membrane junctional regions between adjacent fibres have also been tentatively identified⁶. We now describe electrophysiological investigations of cellular communication in the cephalopod lens, which show marked differences in the intercellular electrical coupling within the vertebrate (amphibian) and cephalopod lens.

Table 2 Timing of the initiation of the first HR in relation to the first sperm 'stop'

Time (s)	Eggs	Time (s)	Eggs
-5-0	11	0-5	10
-10--5	5	5-10	6
-15--10	1	10-15	2
----15	0	15-20	2
		20----	4

The time when the flagellar motion of the first sperm ceased was taken as zero time.

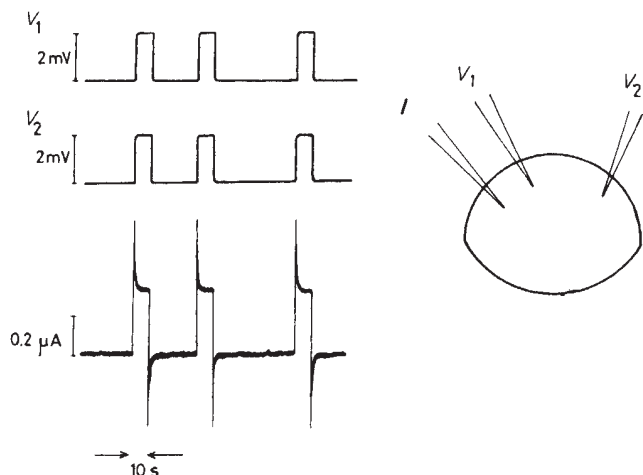


Fig. 1 Voltage-clamp responses at two positions (V_1 and V_2) within the cortex of the perfused lens of the frog *Rana pipiens* (for dissection and solution details, see ref. 9). The experimental arrangement allowed for two separate voltage-measuring circuits. The first electrode (V_1) measured the lens transmembrane voltage with respect to a reference in the bath and was coupled by a voltage-clamp circuit to the current-injection microelectrode (I) so that the lens could be clamped at any given voltage (see ref. 9 for further details). The voltage was also measured by an independent differential amplifier circuit between a third internal microelectrode (V_2) and a separate reference electrode in the solution flowing round the lens. The upper trace shows 2-mV command pulses superimposed on a resting voltage of -75 mV. The middle trace gives the response of the independent electrode where the resting baseline was -74 mV. V_2 in this case was placed one half of a lens diameter from V_1 . The clamp current (I) consists of a sharp spike, indicating current flow through the large membrane capacitance¹⁰ followed by a decline to a plateau value when current is flowing through the pure resistive components of the lens membranes. The voltage and current responses were found to be independent of the relative positions of the three widely-spaced electrodes.

Insertion of three microelectrodes into the frog lens does not significantly impair its membrane characteristics as stable, high (mean $\pm$ s.e.m. = 77.8 ± 0.7 mV from 29 lenses) resting potentials were obtained during these experiments. The voltage recorded at electrode V_1 was always within 1–2 mV of that recorded at V_2 . When V_1 was clamped at the resting voltage (Fig. 1), V_2 also remained clamped and when a 2-mV pulse (either hyperpolarizing or depolarizing) was superimposed on the resting voltage (V_1), a very similar change in baseline was also recorded at V_2 . The coupling ratio (V_1/V_2) obtained in four such experiments was 0.98 ± 0.1 .

However, the voltage-clamp responses obtained from three similar internal electrode experiments on the perfused lens of the cephalopod *Sepiolo atlantica*, depended on the relative positions of the current and voltage electrodes (Fig. 2). When the current and command voltage electrodes were both placed in the anterior segment and the independent voltage electrode in the posterior, a 2-mV command pulse was not accompanied by a 2-mV response at the posterior electrode (Fig. 2a); the coupling ratio defined by V_p/V_a was in fact only 0.06. These responses were mirrored when the command and current electrodes were

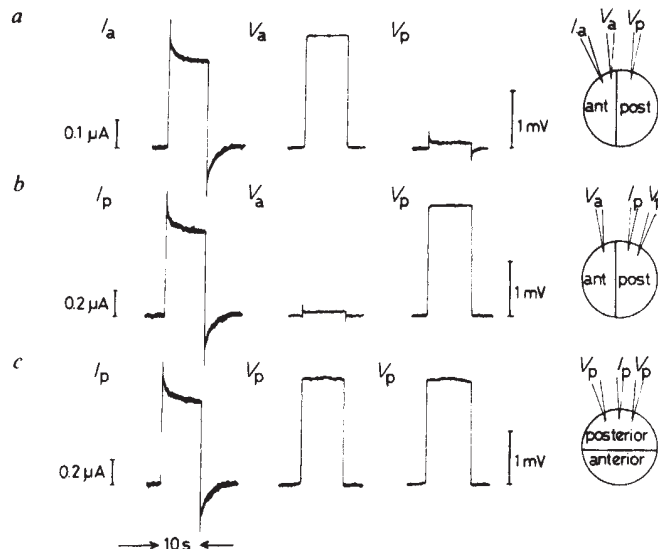


Fig. 2 Voltage-clamp responses at the anterior and posterior segments of the perfused cephalopod lens (for details of dissection methods and solution composition, see ref. 6). The command voltage microelectrode and current injection electrode were always placed in the same segment of the lens and the shape of the voltage and current signals obtained in response to a command pulse of 2 mV were very similar to those obtained from the vertebrate lens (Fig. 1). However, the magnitude and time course of the response measured by the independent voltage circuit depended on whether the command and independent electrode were located in the same lens segment. When they were in different segments (as in a and b), the independent signal was greatly reduced in amplitude and its shape followed that of the clamp current rather than the command voltage. When the two voltage electrodes were in the same segment (as in c), the amplitudes and shapes of the responses were very similar. Note the difference in vertical scale magnitudes of the clamp currents in a and b, which is due to the fact that the resistance of the anterior segment is greater than that of the posterior (Table 1).

both in the posterior segment, with the independent electrode in the anterior (Fig. 2b); the coupling ratio (V_a/V_p in this case) was again 0.06 and the mean value obtained from eight such experiments was 0.072 ± 0.015 . When the three electrodes were placed in the same segment (either anterior or posterior), the command pulse was followed precisely by the independent electrode (Fig. 2c). Hence cells within one segment are closely coupled, but there is little communication between segments.

The resting voltages recorded in the two segments were very similar (Table 1), but the resistances were quite different. However, the surface area of the anterior segment in the *Sepiolo* lens is about half that of the posterior (our unpublished observations), thus when the data are expressed as specific resistance ($\text{k}\Omega \text{ cm}^2$) the two values are almost equal. This suggests that although the membranes of the fibre cells in the two segments have a different origin and although they are not coupled, their membrane permeability properties may be quite similar. The specific resistances of the *Sepiolo* segments are both considerably smaller than the specific resistance of the frog lens, which Delamere and Duncan⁶, ascribed to the very high potassium permeability of the *Sepiolo* lens membranes.

Table 1 Accumulated potential and resistance data for the amphibian (*Rana pipiens*) and cephalopod (*Sepiolo atlantica*) lenses

	Potential (mV)	Resistance ($\text{k}\Omega$)	Surface area ($\times 10^{-2} \text{ cm}^2$)	Specific resistance ($\text{k}\Omega \text{ cm}^2$)
<i>Rana</i> lens	-77.8 ± 0.7 (29)	10.9 ± 0.48	33.9 ± 1.17	3.69 ± 0.144
<i>Sepiolo</i> lens anterior	-62.5 ± 1.4 (15)	6.34 ± 0.49	6.9 ± 0.3	0.44 ± 0.026
posterior	-62.2 ± 2.0 (9)	3.49 ± 0.87	15.3 ± 0.59	0.53 ± 0.07

Data are mean $\pm$ s.e.m. and values in parentheses represent the number of lenses in each case.

The fact that the two segments of the cephalopod lens are not closely coupled (Fig. 2a, b) poses interesting problems for the regulation of growth and raises a further issue concerning the crystallin protein composition of the two halves. As they are of different ontogenetic origin, and because they remain functionally separate throughout development, it is very likely that they are composed of different protein species. Although the proteins of cephalopod lenses have been investigated in some detail⁷ and the nuclear and cortical components seem to be different,⁸ no studies have been made of the two halves separately.

The cephalopod lens thus provides a unique opportunity for comparative evolutionary and developmental studies. Not only has it developed one mechanism in parallel with the vertebrate lens for producing a highly refractile, transparent focusing system (that is, by close-packed array of fibre cells of surface ectodermal origin), but it has also evolved a second and apparently independent mechanism for producing crystallins and fibre cells from a second cell type (vesicular epithelium).

Received 2 December 1980; accepted 29 January 1981.

- Philipson, B. T., Hanninen, L. & Balazs, E. A. *Expl Eye Res.* **21**, 205–219 (1975).
- Kusak, J., Maisel, H. & Harding, C. V. *Expl Eye Res.* **27**, 495–498 (1978).
- Duncan, G. *Expl Eye Res.* **8**, 406–414 (1969).
- Duncan, G. in *The Eye* Vol. 5 (eds Davson, H. & Graham, L. T.) 357–398 (Academic, London 1974).
- Duke-Elder, S. *System of Ophthalmology* Vol. 1 (Kimpton, London, 1958).
- Delamere, N. A. & Duncan, G. *J. Physiol., Lond.* **272**, 167–186 (1977).
- Clayton, R. M. in *The Eye* Vol. 5 (eds Davson, H. & Graham, L. T.) 399–494 (Academic, London, 1974).
- Dohrn, A. *Expl Eye Res.* **9**, 297–299 (1970).
- Jacob, T. J. C. & Duncan, G. *Expl Eye Res.* **31**, 505–512 (1980).
- Duncan, G., Patmore, L. & Pynsent, P. B. *J. Physiol., Lond.* **312** (in the press).

Rhodopsin-like photosensitivity of isolated chicken pineal gland

Takeo Deguchi

Department of Medical Chemistry, Tokyo Metropolitan Institute for Neurosciences, 2–6 Musashidai, Fuchu-city, Tokyo 183, Japan

The pineal gland of lower vertebrates such as fish, amphibians and reptiles is primarily a photoreceptor organ and its electrical activity changes in response to environmental lighting^{1–3}, whereas the mammalian pineal gland is a secretory organ and no longer responds to direct illumination. It is uncertain whether the avian pineal gland retains a photoreceptive capability^{4–8}, and also whether the gland is a photosensor involved in the photo-periodic control of reproductive systems and circadian rhythms in birds^{9–14}. Recently, I and other workers have shown that the circadian rhythm of serotonin *N*-acetyltransferase activity persisted in organ cultures of chicken pineals, indicating the presence of an endogenous circadian oscillator in the gland^{15–19}. Also, the observation that direct illumination of cultured chicken pineals suppressed the night-time increase of *N*-acetyltransferase activity^{16,20} suggests that the gland contains a photoreceptor. I report here that the action spectrum of the photosensitivity of the isolated chicken pineal resembles the absorption spectrum of rhodopsin.

Chickens (White Leghorn cockerels) were maintained in our animal facility with light from 7 a.m. to 7 p.m. for 14–18 days after hatching. The chickens were killed by decapitation at 6 p.m. and five pineals were placed on nylon mesh in a Falcon plastic organ culture dish (no. 3037) containing 1.5 ml of BGJb medium without phenol red. They were cultured as described elsewhere¹⁶ from 7 p.m. to 1 a.m. under a monochromatic and energy-calibrated photic stimulus. As controls, a group of pineals was cultured in darkness (dark control) and another

Table 1 Effect of depolarizing and hyperpolarizing agents on *N*-acetyltransferase rhythm in organ cultures of chicken pineal gland

Addition	<i>N</i> -acetyltransferase (pmol per 10 min per pineal)	
	Light	Dark
None	152 ± 12	997 ± 50
55 mM KCl	434 ± 58	990 ± 131
Nonactin (20 µM)	115 ± 13	107 ± 15
Valinomycin (4 µM)	129 ± 10	135 ± 11
Before culture	98 ± 6	

Pineals were organ-cultured from 7 p.m. to 1 a.m. in darkness or under 70 µW cm⁻² of white light in the presence of the agents indicated. There were five samples in each group and the results are the means ± s.e.m.

group under 70 µW cm⁻² of white light from a xenon arc lamp (light control).

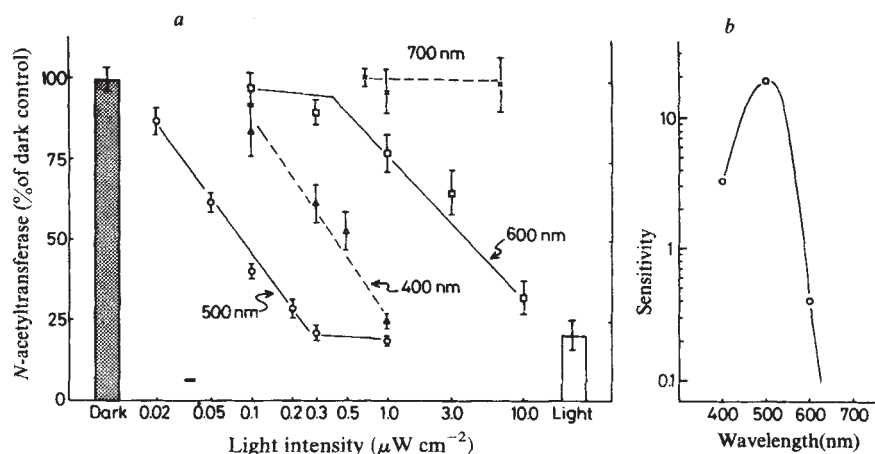
When pineals were cultured from 7 p.m. to 1 a.m. in darkness, *N*-acetyltransferase activity increased from 75–120 to 750–1,300 pmol per 10 min per pineal. When they were cultured under continuous illumination, the enzyme activity increased only to 150–250 pmol per 10 min per pineal. Culture under various intensities of 400, 500, 600 and 700 nm wavelengths (Fig. 1a) showed 500 nm to be the most effective wavelength for inhibiting *N*-acetyltransferase activity. At this wavelength there was a significant decrease of the enzyme activity with an intensity of 0.02 µW cm⁻² ($P < 0.05$), and maximum inhibition with 0.3 µW cm⁻², which was equivalent to that caused by 70 µW cm⁻² of white light. With the wavelength of 400 nm maximum inhibition occurred at 1 µW cm⁻², while 600 nm was far less effective, showing no significant inhibition with 0.3 µW cm⁻² and a submaximum inhibition with 10 µW cm⁻². The wavelength of 700 nm showed no effect on enzyme activity with 7 µW cm⁻² or less. When the reciprocals of the light intensities producing 50% inhibition of the night-time increase of *N*-acetyltransferase activity were plotted as sensitivities on the ordinate in a logarithmic scale (Fig. 1b), the sensitivity curve resembled the absorption spectrum of rhodopsin²¹. On examining the effects of wavelengths from 400 to 800 nm at 0.3 µW cm⁻², maximum inhibition occurred at 500 nm and a submaximum inhibition at other wavelengths (Fig. 2). The most effective wavelengths were 500 and 520 nm, and there was a sharp cutoff of the inhibitory effect at longer wavelengths, with a gradual decrease at shorter wavelengths.

A high enough concentration of KCl (55 mM) to depolarize the cell membrane partially restored the light-induced suppression, but did not enhance the dark-time increase of *N*-acetyltransferase activity (Table 1). On the other hand, the dark-time increase of *N*-acetyltransferase activity was completely inhibited by potassium ionophores (nonactin or valinomycin), which hyperpolarize the cell membrane. I have previously shown that cocaine, which prevents depolarization of neural cells, and catecholamines, which induce hyperpolarization in the rat pinealocyte, block the night-time elevation of *N*-acetyltransferase activity²².

The threshold light intensity of the chicken pineal is in the range of the threshold intensities of several photosensitive systems²³. As dispersed cells responded to a photic stimulation in the same manner as organ culture²⁴, it seems reasonable to speculate that a rhodopsin-like photosensitive substance is incorporated in the cell membrane of pinealocytes. Electron microscopic studies of avian pineals revealed cells with whorls of lamellae and a modified structure, which some investigators consider photosensory^{6,7} but others believe to be abortive or vestigial photoreceptor cells⁸. Such a structure might participate in the regulation of *N*-acetyltransferase activity.

It is firmly established that the membrane potential of retinal photoreceptor cells of vertebrates and of frog pineals are depolarized in darkness and hyperpolarized by photic stimulation^{3,25}. It is therefore conceivable that direct illumination of the

Fig. 1 Effect of various intensities of monochromatic light on the night-time increase of *N*-acetyltransferase activity in organ cultures of chicken pineal. A monochromatic stimulator (model TD-3, Sanso Seisaku) with a xenon arc lamp (XW-500, 500 W) was used. White light from the xenon lamp was divided into three channels and the light energy was calibrated by using an optical wedge and neutral density filter. Monochromatic light was obtained by interference filters (half width <11 nm) and the beam was transmitted into a CO₂ incubator through a quartz window. The stimulating light was reflected downwards with a mirror and projected on to a culture dish. The intensity of light was measured at the level of the dish by use of a radiometer (model 470D, Sanso Seisaku) and expressed as $\mu\text{W cm}^{-2}$. *a*, Pineals were cultured from 7 p.m. to 1 a.m. under the intensities of monochromatic light indicated. *N*-acetyltransferase activity is shown as per cent of dark control, which was run in each experiment. Points and bars show means and s.e.m. of values of five samples. *b*, Reciprocals of the intensities that produced 50% inhibition are plotted as sensitivities on the ordinate in a logarithmic scale.



chicken pineal produces a slow, long-lasting hyperpolarization of the pinealocyte membrane and prevents the nocturnal increase of *N*-acetyltransferase activity. Such speculation is supported by the present finding that depolarizing agents restored the light-suppression of *N*-acetyltransferase activity and hyperpolarizing agents prevented the night-time increase of the enzyme activity.

Kato *et al.* reported that red, but not green, radioluminous paint applied to the skin above the pineal gland of Japanese quail promoted testicular development¹⁴. Woodard *et al.* also showed that in Japanese quail red light accelerated sexual maturity more effectively than green light²⁶. In contrast, Menaker showed that green light (0.1 lx) was effective in entraining the circadian rhythm of perching activity in blinded sparrows²⁷. In my study, green light was most effective in inhibiting the *N*-acetyltransferase rhythm. The discrepancy between these observations is unclear. However, in the present study pineals were directly illuminated, whereas in other studies animals were placed under a specified wavelength of light. It is possible that shorter wavelengths were absorbed by skin, skull, blood and other tissues before reaching the pineals. Alternatively, the hypothalamic, but not pineal, photoreceptor may sense longer wavelengths of light, thus affecting photoreproductive function in living birds.

Menaker and his associates have demonstrated that in house sparrows the pineal gland functions as a biological clock that

controls the circadian rhythms of locomotor activity and body temperature²⁸. The gland presumably regulates such functions by secretion of melatonin, a pineal hormone^{29,30}. The activity of serotonin *N*-acetyltransferase, the key enzyme in the synthesis of melatonin³¹⁻³³, shows a circadian rhythm and responds to direct illumination in the isolated chicken pineal glands, as reported here. Recently, Takahashi *et al.*³⁴ have shown that the synthesis and release of melatonin from cultured chicken pineal glands also exhibit a circadian change and respond to environmental lighting, suggesting that the light perception of the avian pineal gland, by affecting the synthesis of melatonin, has a physiological role in the regulation of functions and behaviour in birds.

This research was in part supported by a Grand-in-Aid for Special Project Research on Animal Behaviors and by a Grant-in-Aid for Scientific Research (548121) from the Ministry of Education of Japan.

Received 24 October 1980; accepted 11 February 1981.

- Wurtman, R. J., Axelrod, J. & Kelly, D. E. in *The Pineal*, 9-39 (Academic, New York, 1968).
- Dodt, E. in *Handbook of Sensory Physiology* Vol. VII/3B (ed. Jung, R.) 113-140 (Springer, Berlin, 1973).
- Morita, Y. in *Brain-Endocrine Interaction* (eds Knigge, K. M., Scott, D. E., Kobayashi, H. & Ishii, S.) 346-387 (Karger, Basel, 1975).
- Morita, Y. *Experientia* **22**, 402-403 (1966).
- Ralph, C. L. & Dawson, D. C. *Experientia* **24**, 147 (1968).
- Quay, W. B., Renzoni, A. & Eakin, R. M. *Riv. Biol.* **61**, 371-393 (1968).
- Bischoff, M. B. *J. ultrastruct. Res.* **28**, 16-26 (1969).
- Oksche, A. & Vaupel-Von Harnack, M. Z. *Zellforsch.* **69**, 41-60 (1966).
- Harrison, P. C. & Becker, W. C. *Proc. Soc. exp. Biol. Med.* **132**, 161-164 (1969).
- Homma, K. & Sakakibara, Y. in *Biochronometry* (ed. Menaker, M.) 333-341 (National Academy of Science, Washington DC, 1971).
- Menaker, M., Roberts, R., Elliott, J. & Underwood, H. *Proc. natn. Acad. Sci. U.S.A.* **67**, 320-325 (1970).
- Oksche, A., Kirschstein, H., Kobayashi, H. & Farner, D. A. Z. *Zellforsch. mikrosk. Anat.* **24**, 247-274 (1972).
- Hisano, N., Cardinali, D. P., Rosner, J. M., Nagle, C. A. & Tramezzani, J. H. *Endocrinology* **91**, 1318-1322 (1972).
- Kato, M., Kato, T. & Oishi, T. *Proc. Jap. Acad.* **43**, 220-223 (1967).
- Deguchi, T. in *Integrative Control Functions of the Brain* Vol. 1 (eds Ito, M., Tsukahara, N., Kubota, K. & Yagi, K.) 345-347 (Kodansha, Tokyo/Elsevier, Amsterdam, 1978).
- Deguchi, T. *Science* **203**, 1245-1247 (1979).
- Binkley, S. A., Riebelman, J. B. & Reilly, K. B. *Science* **202**, 1198-1201 (1978).
- Kasal, C. A., Menaker, M. & Perez-Polo, J. R. *Science* **203**, 656-658 (1979).
- Wainwright, S. D. & Wainwright, L. K. *Can. J. Biochem.* **57**, 700-709 (1979).
- Wainwright, S. D. & Wainwright, L. K. *J. Neurochem.* **35**, 451-457 (1980).
- Wald, G. & Brown, P. K. *Nature* **177**, 174-176 (1956).
- Deguchi, T. *J. Neurochem.* **33**, 45-51 (1979).
- Andresen, M. C. & Brown, A. M. *J. Physiol., Lond.* **287**, 267-282 (1979).
- Deguchi, T. *Nature* **282**, 94-96 (1979).
- Tomita, T. *Q. Rev. Biophys.* **3**, 179-222 (1970).
- Woodard, A. E., Moore, J. A. & Wilson, W. O. *Poultry Sci.* **47**, 1733-1737 (1967).
- Menaker, M. *Proc. natn. Acad. Sci. U.S.A.* **59**, 414-421 (1968).
- Menaker, M., Takahashi, J. S. & Eskin, A. A. *Rev. Physiol.* **40**, 501-526 (1978).
- Gwinner, E. & Benzinger, I. J. *comp. Physiol.* **127**, 209-213 (1978).
- Hendel, R. C. & Turek, F. W. *Physiol. Behav.* **21**, 275-278 (1978).
- Keitel, D. C. & Weller, J. L. *Science* **169**, 1093-1095 (1970).
- Deguchi, T. & Axelrod, J. *Proc. natn. Acad. Sci. U.S.A.* **69**, 3547-3550 (1972).
- Axelrod, J. *Science* **184**, 1341-1348 (1974).
- Takahashi, J. S., Hamm, H. & Menaker, M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2319-2322 (1980).

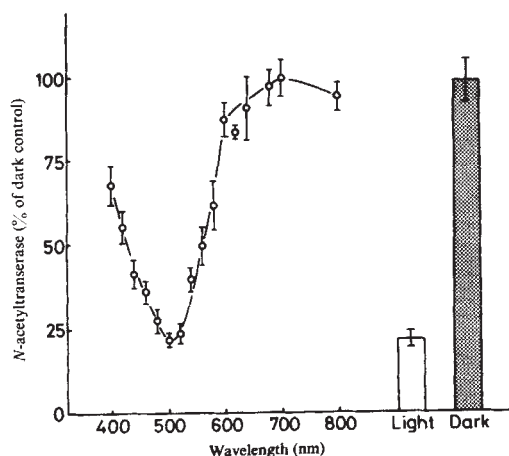


Fig 2 Effect of various wavelengths on the *N*-acetyltransferase activity in organ culture. Pineals were cultured from 7 p.m. to 1 a.m. under an intensity of $0.3 \mu\text{W cm}^{-2}$ of the monochromatic light indicated. The results are shown as per cent of dark control. Points and bars indicate means and s.e.m. of five samples.

Functionally significant central-pair rotation in a primitive eukaryotic flagellum

Charlotte K. Omoto*† & George B. Witman*

* Department of Biology, Princeton University, Princeton, New Jersey 08544, USA

† Biology Division, California Institute of Technology, Pasadena, California 91125, USA

There is now considerable evidence that the basis for ciliary and flagellar movement is an active sliding between peripheral doublet microtubules which, when resisted by structures within the axoneme, leads to axonemal bend formation¹⁻⁴. In contrast, relatively little is known about the control mechanisms which coordinate the interdoublet sliding and axonemal bending to produce the effective motion observed in various cilia and flagella⁵. One component of the axoneme which may be involved in this control is the central pair of microtubules^{6,7}. To learn more about the action of the central pair, we have studied the tiny uniflagellate marine alga, *Micromonas pusilla*. The central tubules of the *M. pusilla* flagellum extend for several micrometres beyond the termination of the peripheral doublets⁸, thus permitting direct observation of the central pair during flagellar movement. Our findings, reported here, indicate that in living *M. pusilla* the central pair of microtubules undergoes continuous rotation in one direction. This rotation provides the motive force for the cell.

Our observations on the ultrastructure of the flagellum of *M. pusilla* (previously named *Chromulina pusilla*⁹) agree well with those of Manton⁸. The flagellum is ~5 µm long and has the appearance of a short thick stub terminated by a long thin fibre (Fig. 1a). The thicker proximal portion is about 0.2 µm in diameter and contains the normal '9+2' array of microtubules, but is less than 1 µm long (Fig. 1e). The thinner distal portion is ~70 nm in diameter and contains only the central pair of microtubules (Fig. 1b), which are ~5 µm in length and extend for ~4 µm beyond the termination of the peripheral doublets. For most of this distal portion, the flagellar membrane closely surrounds the central pair. Figure 1b and c show the region of transition between the '9+2' flagellum and the thin '0+2' portion containing only the central tubules. At their proximal end, the central tubules terminate just above the basal body (Fig. 1f), as in other cilia and flagella^{10,11}.

The flagella of living *M. pusilla* can be observed by dark-field light microscopy. In free-swimming cells, the distal '0+2' portion takes on a helical configuration and seems to rotate, pushing the cells through the medium like a propeller. This portion of the flagellum also appears to rotate in cells which are stuck to the glass slide by their cell bodies (Fig. 2). The short '9+2' region is usually obscured by glare from the cell body, but in fortuitous views it is seen to wave back and forth, as would be expected for a very short flagellum undergoing a normal beat cycle. The '0+2' portion of a non-moving flagellum retains its helical configuration. By focusing through such a flagellum, we determined that the helix is left-handed.

One cannot be sure from the above observations alone that the movement of the '0+2' portion of the flagellum is due to actual rotation as opposed to propagation of helical waves along a non-rotating flagellum. However, these two possibilities can be distinguished by anchoring the cell to the glass slide by its flagellum and observing the resulting motion of the cell body. When *M. pusilla* are placed on silicone-coated slides, many cells stick to the slide by one end of the cell body and spin rapidly around the point of attachment, as previously described by Manton and Parke⁹. However, some cells stick to the slide by the flagellum; when this occurs, the cell body rotates (Figs 3, 4). In

Fig. 3, the rotational axis is almost perpendicular to the slide, so that the rotation looks like that of a wheel viewed from the side. In Fig. 4, the rotational axis is nearly parallel to the glass slide and rotation is inferred by the changes seen in the shape and the markings of the cell. The position and curvature of the flagellum of this cell remain constant throughout the 360° rotation of the cell body, indicating that the flagellum is stuck to the slide and immobilized over most of its length. In all such cells so far observed, the cell bodies were rotating clockwise as viewed from the flagellum. This would correspond to an anticlockwise rotation of the '0+2' portion, looking from tip to base. These observations indicate that the distal portion of the *M. pusilla* flagellum, and by inference, the pair of central tubules therein,

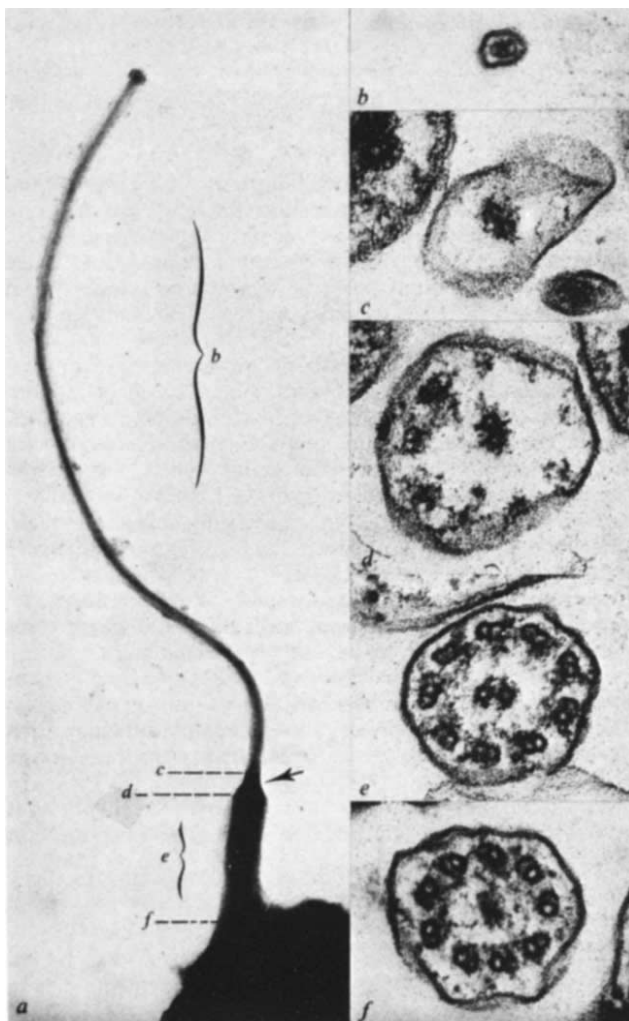


Fig. 1 Structure of *M. pusilla* flagellum. *a*, Electron micrograph of a negatively stained *M. pusilla*. The distal portion of the flagellum beyond the arrow is about 70 nm in diameter and is too thin to contain the usual '9+2' axoneme. The dotted lines and brackets labelled *b-f* indicate the approximate levels of the corresponding cross-sections. *M. pusilla* was grown under continuous light on Ergschieber's medium²⁰. A drop of the cell culture was placed on a Formvar-coated grid and stained with 2% uranyl acetate. $\times 15,275$. *b*, Cross-section through distal portion of the flagellum where only the central tubules remain. Note that the membrane closely surrounds the two tubules, which seem to be closer to each other than in the usual '9+2' structure. *c*, Cross-section through the flagellum just distal to the point where the peripheral tubules have terminated. *d*, Cross-section where some of the peripheral tubules have terminated. *e*, Cross-section through the proximal portion of the flagellum, showing the typical '9+2' axoneme. *f*, Cross-section at the base of the flagellum, showing the axoneme just proximal to the origin of the central tubules. *b-f*, $\times 83,800$. *M. pusilla* was 'instantaneously fixed'¹² in glutaraldehyde-OsO₄, embedded in agar, then dehydrated and embedded in Spurr's plastic²¹. Sections were stained with uranyl acetate and Reynold's lead citrate²².

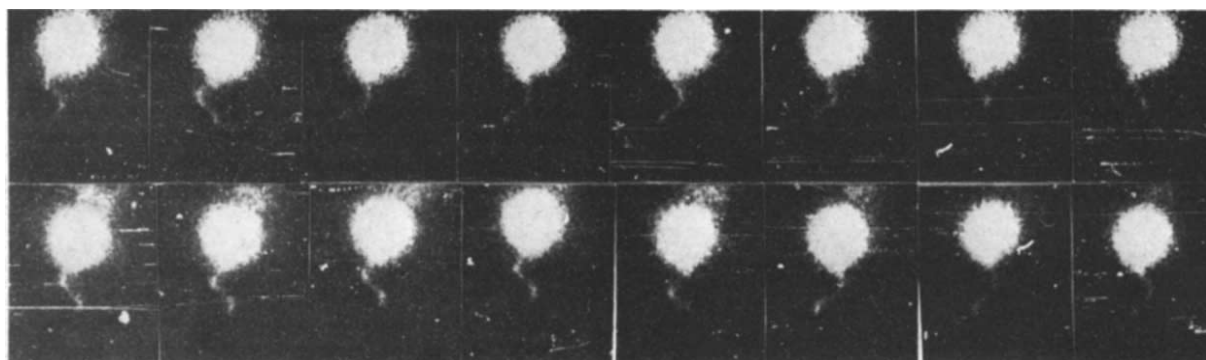


Fig. 2 Sequence from 16-mm cine film of *M. pusilla*. The cell is stuck to the slide and the flagellum seems to rotate. Most of the flagellum seen is the '0+2' region, the '9+2' region being too short to be visible much beyond the glare of the cell body. Dark-field optics⁶; consecutive frames progress from left to right. 64 frames per s. $\times 2,000$.

rotates. The evidence does not rule out the possibility that the whole flagellum, including the '9+2' region, rotates, but this seems highly unlikely.

The anticlockwise rotation together with the left-handed twist of the flagellum are consistent with the observed progression of the cell through the medium, with the flagellum pushing the cell. Interestingly, the direction of central-pair rotation in *M. pusilla* is the same as that reported for the central pair in *Paramecium*^{7,12}.

The membrane which surrounds the central microtubules probably rotates with the central pair; this is especially evident in a rotating cell which is anchored to the slide by its flagellum, but even in free-swimming cells connections between the central pair and the membrane might cause the membrane to move with the underlying tubules. Such a movement would result in a shear zone in the flagellar membrane at some point along the flagellum, probably at the base of the '0+2' portion. We have observed no discontinuity in the cell or flagellar membrane of *M. pusilla*, so the surface shear must be readily absorbed by the fluid membrane. A similar situation occurs in the plasma membrane of some devescovid flagellates, where one portion of the cell

surface continuously rotates relative to the rest of the cell surface¹³.

We do not know what causes the central pair to rotate. One possibility is that the rotation is a passive response to three-dimensional bending of the '9+2' portion of the flagellum. Alternatively, rotation may be the result of interactions between the radial spokes and the central sheath in the '9+2' region, or might be caused by a rotary motor in the axosome where the central tubules originate. Further studies will be necessary to determine the basis for the rotation.

Central-pair rotation has now been demonstrated for *M. pusilla* flagella, and implicated in *Paramecium* cilia^{7,12}, *Opalina* cilia¹⁴ and *Synura* flagella¹⁵. Central-pair rotation may therefore occur in the cilia and flagella of many organisms. In *M. pusilla*, rotation of the central pair propels the cell through its medium. The function of central-pair rotation in the normal '9+2' axoneme is unknown, but several observations suggest that the central microtubules are involved in the control of ciliary and flagellar movement. Flagella lacking the central tubules are paralysed, although the ability of the dynein arms to cause sliding is apparently normal¹⁶. The central tubule-central

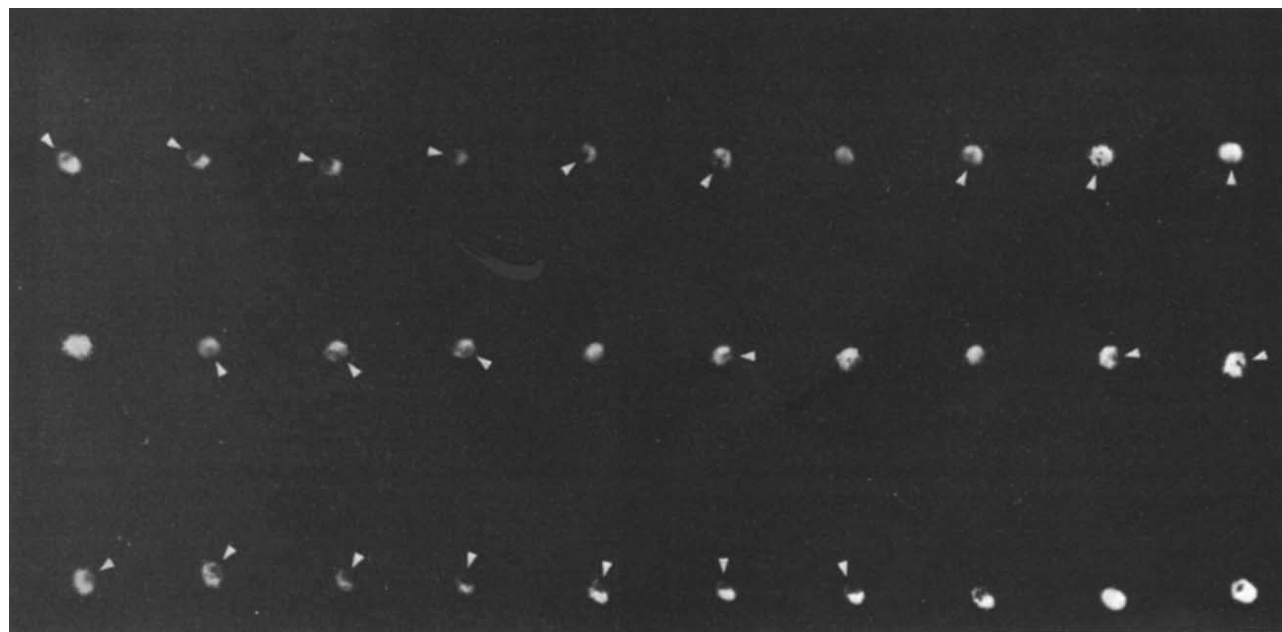


Fig. 3 Film sequence showing a cell stuck by its flagellum (visible in the original cine film) to a silicone-coated slide and the cell body rotating. The slide was coated with silicone by wiping it with silicone-treated lens paper (Sight Savers, Dow-Corning). The sequence was printed to enhance the visibility of the dark spot (white arrowheads) on the cell body. The cell-body rotation can be followed using the dark spot as a marker. The cell is rotating anticlockwise. Consecutive frames run from left to right beginning in the upper left-hand corner. Dark-field optics, 64 frames per s. $\times 2,880$.

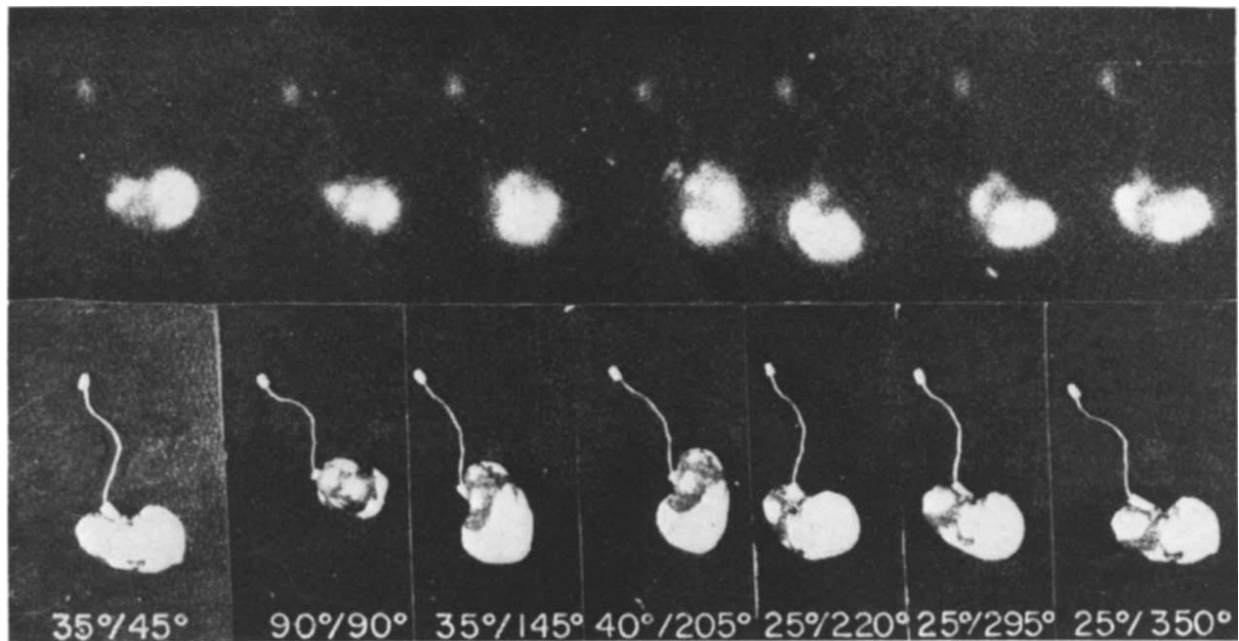


Fig. 4 Top row: dark-field light micrographs of *M. pusilla*. The cell is attached to a silicone-coated slide by its flagellum and the asymmetric cell body is rotating. Recorded on continuously moving film with stroboscopic illumination at 60 Hz. $\times 3,200$. Second row: photos of a styrofoam and wire model. The wire, representing the '0 + 2' portion, is held in place and the angle of insertion of the wire and the cell body orientation are changed and photographed. The numbers below each image indicate the angle of insertion relative to the cell surface, and the angle through which the cell body has rotated relative to the surface to which the flagellum is attached, respectively.

sheath complex is asymmetrical in several organisms, and this asymmetry may be related to the formation of asymmetrical axonemal bends^{6,16}. The orientation within the axoneme of the central pair of microtubules is correlated to beat direction in a variety of cilia and flagella^{14,17,18}; moreover, the orientation of the central pair has been reported to change along the length of the axoneme^{12,19}, with change in beat direction¹⁴, and during the beat cycle^{7,12,15}. Thus, it has been proposed that the central microtubules undergo a systematic movement correlated with the beat cycle and that this movement is involved in regulating the motion of the axoneme, possibly by determining which doublets undergo active sliding at any particular phase in the beat cycle⁷. Regulation of dynein arm activity around the axoneme would be particularly important in cilia and flagella having a three-dimensional waveform.

The cultures of *M. pusilla* were provided by the Culture Centre of Algae and Protozoa (Cambridge, UK) and the Culture Collection of Algae (University of Texas, Austin). We thank Jana Olson Kiefer for providing the growth medium, Nick Minervini for assistance with photography and electron microscopy, Gary Gorbisky for his suggestion of using 'Sight Savers' for coating slides and Drs Ching Kung, Howard Berg and Charles J. Brokaw for helpful criticism and encouragement. Figure 4 was made possible by the invaluable assistance of Dr Brokaw at the California Institute of Technology, Kerckhoff Marine Laboratory. This work was supported in part by NIH grants GM 21586 to G.B.W., GM 07445 to C.K.O. and GM 18711 to C.J.B.

Received 8 October; accepted 2 February 1981.

1. Satir, P. *J. Cell Biol.* **39**, 77-94 (1968).
2. Summers, K. E. & Gibbons, I. R. *Proc. natn. Acad. Sci. U.S.A.* **68**, 3092-3096 (1971).
3. Gibbons, B. H. & Gibbons, I. R. *Biochem. biophys. Res. Commun.* **73**, 1-6 (1976).
4. Shingyogi, C., Murakami, A. & Takahashi, K. *Nature* **265**, 269-270 (1977).
5. Brokaw, C. J. & Gibbons, I. R. in *Swimming and Flying in Nature* (eds Wu, T. Y.-T. et al.) 89-126 (Plenum, New York, 1975).
6. Bessen, M. et al. *J. Cell Biol.* **86**, 446-455 (1980).
7. Omoto, C. K. & Kung, C. *Nature* **279**, 532-534 (1979).
8. Manton, I. *J. mar. Biol. Ass. U.K.* **38**, 319-333 (1959).
9. Manton, I. & Parke, M. *J. mar. Biol. Ass. U.K.* **39**, 275-298 (1960).
10. Ringo, D. L. *J. Cell Biol.* **35**, 543-571 (1967).
11. Dute, R. & Kung, C. *J. Cell Biol.* **78**, 451-464 (1978).
12. Omoto, C. K. & Kung, C. *J. Cell Biol.* **87**, 33-46 (1980).
13. Tamm, S. L. & Tamm, S. J. *Cell Sci.* **20**, 619-639 (1976).
14. Tamm, S. L. & Horridge, G. A. *Proc. R. Soc. B* **175**, 219-233 (1970).
15. Jarosch, R. & Fuchs, B. *Protoplasma* **85**, 285-290 (1975).
16. Witman, G. B. et al. *J. Cell Biol.* **76**, 729-747 (1978).

17. Afzelius, B. A. *J. Biophys. Biochem. Cytol.* **9**, 383-394 (1961).
18. Gibbons, I. R. *J. Biophys. Biochem. Cytol.* **11**, 179-205 (1961).
19. Gibbons, I. R. in *Molecules & Cell Movement* (eds Inoué, S. & Stephens, R. E.) 207-232 (Raven, New York, 1975).
20. Starr, R. C. *J. Phycol.* **14**, Suppl., 47-100 (1978).
21. Spurr, A. R. *J. ultrastruct. Res.* **26**, 31-43 (1969).
22. Reynolds, E. S. *J. Cell Biol.* **17**, 208-212 (1963).

Two phospholipase pools for prostaglandin synthesis in macrophages

Wei Hsueh, Usha Desai, F. Gonzalez-Crussi, Ruth Lamb & Agnes Chu

The Department of Pathology of The Children's Memorial Hospital, Northwestern University Medical School, Chicago, Illinois 60614, USA
The Department of Pathology of University of Connecticut Health Center, Farmington, Connecticut 06032, USA

Macrophages in culture produce prostaglandins in response to a variety of phagocytic and non-phagocytic stimuli¹⁻⁸. As prostaglandins are not stored in cells, and mammalian cells contain very little free arachidonic acid, synthesis and release of prostaglandins depends on the release of the precursor, arachidonic acid, from cell lipids. Many agents that stimulate cell prostaglandin production act by releasing arachidonic acid⁹, presumably by activating phospholipase A₂ (refs 10, 11) or phospholipase C (refs 12, 13), depending on the cell system used. We have shown previously that rabbit alveolar macrophages secrete arachidonic acid as well as prostaglandins in response to phagocytic stimuli^{2,4,6}. This secretion depends not on particle attachment, but rather on interiorization of the particles⁴. Furthermore, the time course of prostaglandin and arachidonic acid secretion does not parallel that of particle engulfment *per se*, but of the release of lysosomal enzymes⁴, indicating that the release of arachidonate and prostaglandin is associated with the latter. We now describe experiments which suggest that these are two independent pools of phospholipases in macrophages, one in the lysosomes and one elsewhere.

To assess the prostaglandin release, we prelabelled the cell lipids of cultured rabbit alveolar macrophages (more than 90% macrophages) with ^{14}C -arachidonic acid (Amersham) as previously described². Analysis of the cell lipids showed a pattern of incorporation similar to that of our previous studies²: approximately 40–50% of the radioactive arachidonic acid was incorporated into cell lipids after 1 h of incubation. Most of the radioactivity (75%) was found in phospholipids, particularly in phosphatidylcholine ($48 \pm 1\%$, $n = 10$) and phosphatidylethanolamine ($19 \pm 1\%$). A small amount also resided in phosphatidylserine and phosphatidylinositol ($8 \pm 1\%$) and neutral lipids ($5 \pm 1\%$). Less than 2% of the radioactivity in cells remained in the free fatty acid fraction.

Stimulation of macrophages with various phagocytic and non-phagocytic stimuli resulted in prostaglandin release. However, only a small amount of arachidonic acid could be detected in the medium in the absence of serum. This is probably due to immediate reincorporation of the unreacted arachidonic acid into phospholipids. To avoid this problem, we used fatty acid-free bovine serum albumin to 'trap' the released fatty acid^{2,14}. Thus, the activity of phospholipase could be quantitatively assessed.

We have previously demonstrated a close association between the release of lysosomal enzyme and that of arachidonic acid and prostaglandin by macrophages, following zymosan phagocytosis^{2,4}. This suggests a cause-effect relationship between these two mechanisms. Prostaglandin synthesis, however, does not regulate lysosomal enzyme release, as inhibition of the former did not affect the latter^{2,15}. As alveolar macrophage lysosomal granules do contain phospholipase A_2 (ref. 16), it is possible that phagosome-lysosome fusion, on releasing acid hydrolases, activates lysosomal phospholipase, thus explaining the close

association between release of arachidonic acid and lysosomal enzyme. To investigate further whether activation of lysosomal enzyme *per se* is sufficient to elicit prostaglandin synthesis, we examined the effect of C5 fragment, a chemotactic agent known to elicit lysosomal enzyme release¹⁷, on arachidonic acid release and prostaglandin synthesis. The effect of C5 fragment was then compared with that of zymosan phagocytosis. As shown in Fig. 1b, both zymosan and C5 fragment resulted in significant increased release of arachidonic acid, up to 4–5 times the control. However, these two agents produced different effects on the coupling efficiency of phospholipase (arachidonic acid release) and cyclooxygenase (prostaglandin synthesis). Figure 1a shows that prostaglandin synthesis was increased to 2.5 times the control value following zymosan stimulation, but only to 1.1 times control value after C5 fragment treatment. Thus, phagocytosis of zymosan resulted in tight coupling between arachidonic acid release and prostaglandin synthesis (arachidonate/prostaglandin ratio = 3, Fig. 1d), whereas C5 fragment, with an arachidonate/prostaglandin ratio of 7.7 (Fig. 1d), produced a disproportionately small prostaglandin synthesis for a large arachidonic acid release. Lysosomal enzyme release remained comparable with both stimulants (approximately three times the control value, Fig. 1c).

Because it is generally accepted that phospholipase, not cyclooxygenase, is the rate-limiting enzyme for prostaglandin synthesis, this dissociation between arachidonic acid release and prostaglandin synthesis by C5 fragment can only be explained as

Fig. 1 Release of ^{14}C -prostaglandin, ^{14}C -arachidonic acid and acid phosphatase by rabbit alveolar macrophages in response to different stimuli. Rabbit alveolar macrophages were collected by lung lavage and plated in Dulbecco's modified Eagle's medium (DMEM) containing 15% fetal calf serum as previously described² (5×10^6 cells per ml per plate). After overnight incubation, the cells were changed to serum-free DMEM. ^{14}C -arachidonic acid (specific activity 55 mCi mmol^{-1} , Amersham) was added to the cultures as the sodium salt at a concentration of $1 \times 10^6 \text{ c.p.m. ml}^{-1}$. The cultures were incubated for 1 h to label the cell lipids. The cells were washed twice with DMEM and changed to DMEM containing 2 mg ml^{-1} fatty acid-free bovine serum albumin. Different stimuli with or without indomethacin were added to the cells and were incubated for 2 h. At the end of incubation, the medium was collected, an aliquot (1/10) was saved for acid phosphatase determination by fluorometric methods as described earlier. The remainder was acidified to pH 3.5, extracted with ethyl acetate and applied to silica gel G plates². The solvent system used for separation of prostaglandins and arachidonic acid was the organic phase from ethyl acetate/water/iso-octane/acetic acid, 110:100:50:20 (ref. 25). Zones corresponding to various prostaglandin standards and 12-HETE (prepared according to Goetzl *et al.*²⁶) were scraped and counted in a liquid scintillation counter. Total prostaglandin release (sum of 6-keto $\text{PGF}_{1\alpha}$ + $\text{PGF}_{2\alpha}$ + PGE_2 + PGD_2 , c.p.m. was calculated (a), the zone corresponding to arachidonic acid standard was also scraped and counted (b) and the acid phosphatase activity in the medium was assayed (c). All values are expressed as percentage of unstimulated control. The intracellular acid phosphatase content of the macrophages treated by various stimuli was determined, and found to be the same as that of unstimulated cells (after correction was made for released amount). d, Ratio of release of arachidonase and prostaglandin. The chemotactic fragment of human C5 (C5 fragment) was isolated from human serum and assayed by a modification of Boyden chamber assay as described previously²⁷. The chemotactic activity was expressed in units of ED_{50} . One ED_{50} of chemotaxis is defined as the amount of C5 fragment required for half-maximal chemotactic response in Boyden chamber assay. Z, zymosan (1 mg ml^{-1}); I, indomethacin ($5 \mu\text{g ml}^{-1}$); C5_{fr} , C5 fragment ($75 \text{ ED}_{50} \text{ ml}^{-1}$); PMA, 10^{-6} M . The numbers within the circles give the numbers of experiments. All values are mean $\pm$ s.e.m.

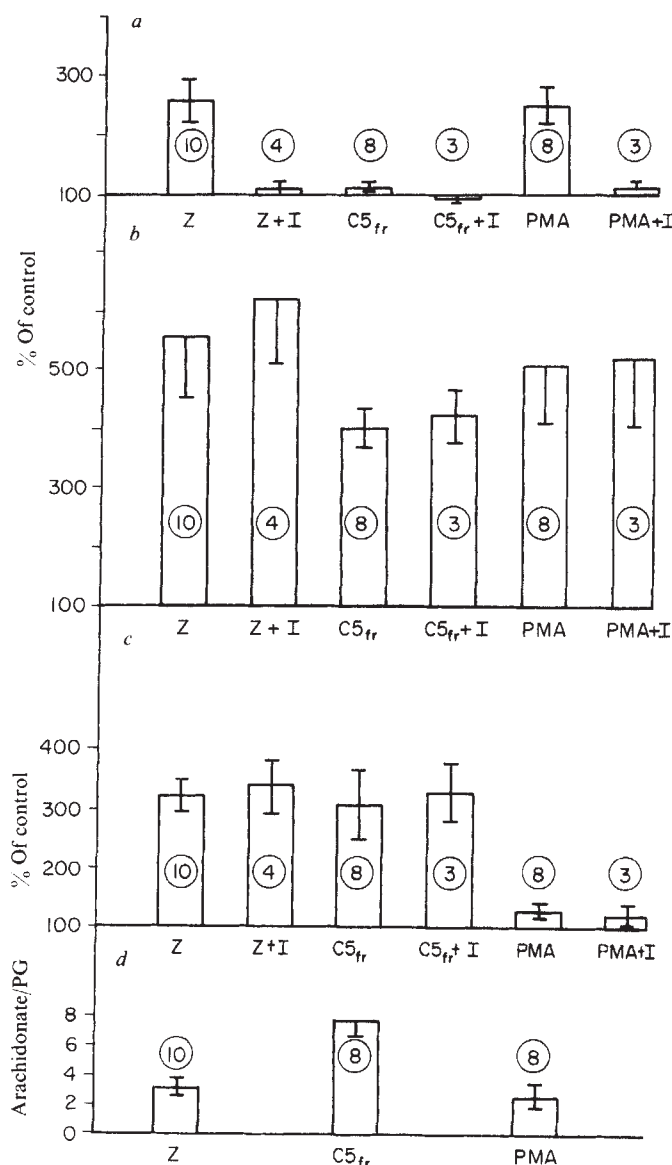


Table 1 Release of arachidonic acid and prostaglandins by alveolar macrophages in response to various stimuli

	Prostaglandins		Arachidonic acid	
	c.p.m.	% Of control	c.p.m.	% Of control
Unstimulated control (<i>n</i> = 9)	750 ± 93	100	898 ± 109	100
Zymosan (1 mg ml ⁻¹) (<i>n</i> = 9)	2,276 ± 283	301 ± 30	9,998 ± 1,326	994 ± 148
PMA (10 ⁻⁶ M) (<i>n</i> = 9)	2,079 ± 294	254 ± 28	7,542 ± 891	736 ± 55
C5 fragment (75 ED ₅₀) (<i>n</i> = 7)	1,008 ± 180	137 ± 21	6,613 ± 909	575 ± 62
Zymosan + PMA (<i>n</i> = 7)	2,942 ± 366	327 ± 45	24,466 ± 3,071	2,681 ± 371
Zymosan + C5 fragment (<i>n</i> = 4)	1,989 ± 265	278 ± 34	10,374 ± 2,059	967 ± 85
C5 fragment + PMA (<i>n</i> = 3)	2,271 ± 618	307 ± 25	38,489 ± 3,152	2,946 ± 384

Prostaglandin (PG) values are the sum of 6-keto PGF_{1α} + PGF_{2α} + PGE₂ + PGD₂(c.p.m.). For methods, see Fig. 1 legend. *n*, Number of experiments, each experiment in duplicate.

follows: (1) there is a switch of arachidonic acid metabolism to the lipoxygenase pathway, (2) C5 fragment inhibits cyclooxygenase, or (3) coupling of phospholipase and cyclooxygenase is prevented, as might occur when a physical distance is interposed between the two enzyme systems within the cell. We can easily rule out (1) as the production of 12-hydroxy-5,8,10,14-eicosatetraenoic acid (12-HETE) and 11-HETE (both co-migrate in solvent system used in this study) by macrophages after C5 fragment stimulation is also very low (120 ± 9% of control compared with 247 ± 31% of control after zymosan stimulation, *n* = 5, mean ± s.e.m.). We cannot entirely rule out the second possibility, as C5 fragment slightly inhibited prostaglandin synthesis when added to zymosan-treated macrophages (Table 1). However, this inhibition is minimal, and cannot account for the discrepancy between arachidonic acid release and prostaglandin synthesis following C5 fragment stimulation. We consider the third explanation a more probable one. Cyclooxygenase activity usually resides in the microsomes, probably in the endoplasmic reticulum^{18,19}. Thus, activation of lysosomal phospholipase by C5 fragment resulted in release of arachidonic acid which remained inaccessible to cyclooxygenase. The phenomena implicated in phagocytosis are different. Zymosan engulfment leads to formation of phagosomes and many intracellular vacuoles²⁰, both of which contain plasma membrane and endoplasmic reticulum, on which cyclooxygenase presumably resides. Fusion of lysosomes with phagosomes and these vacuoles may be expected to result in approximation of activated phospholipase and cyclooxygenase. A consequence of this is more efficient coupling between these two enzyme systems.

Phorbol-myristate acetate (PMA), a membrane perturbant, stimulates the production of prostaglandins in MDCK cells²¹ and fibroblasts²². In rabbit alveolar macrophages, PMA, at the concentration of 10⁻⁶M (non-toxic to cells), resulted in high arachidonate and PG release (5- and 2.5-fold control value, respectively) and a very low lysosomal enzyme release (1.3 times control) (Fig. 1a-d), suggesting the existence of a phospholipase pool independent of lysosomes. Again, the large amount of prostaglandin synthesis is not due to decrease in lipoxygenase pathway, as the production of 11-HETE and 12-HETE is also very high (230 ± 16% of control, *n* = 5, mean ± s.e.m.). Morphological examination of these cells showed numerous

intracellular vacuoles produced by this chemical at a concentration range of 10⁻⁷–10⁻⁶M (Fig. 2a). Both prostaglandin release and vacuole formation became very reduced when the concentration of PMA was reduced to 10⁻⁸M (unpublished data). This observation suggests a close association between vacuole formation and prostaglandin production (Fig. 2a). We propose that: (1) the activated phospholipase resides outside the lysosomes, probably on the plasma membrane; (2) activation takes place during vacuole formation; and (3) the released arachidonic acid is immediately acted on by cyclooxygenase and lipoxygenase at the same site. Such a mechanism correlates with the observed efficient coupling between release of arachidonic acid and release of prostaglandin and HETEs. The small amount of lysosomal enzyme release could be due to fusion of lysosomes with the many cytoplasmic vacuoles. The association of vacuole formation and prostaglandin synthesis in mouse peritoneal macrophages stimulated by 12-O-tetradecanoylphorbol-13-acetate has been reported²⁰ as has the tight coupling of phospholipase and cyclooxygenase following stimulation by phorbol diester and other tumour promoters in cultured canine kidney cells^{23,24}. C5 fragment-treated cells are shown in Fig. 2b for comparison.

The existence of two pools of phospholipase is further supported by the additive effects of zymosan and PMA and of PMA and C5 fragment (Table 1). A combination of PMA (presumably activating membrane phospholipase) with either C5 fragment (presumably activating lysosomal phospholipase) or zymosan (presumably activating lysosomal phospholipase mainly) resulted in supramaximal increase of arachidonic acid release (26.8 times or 29.4 times the control). In contrast,

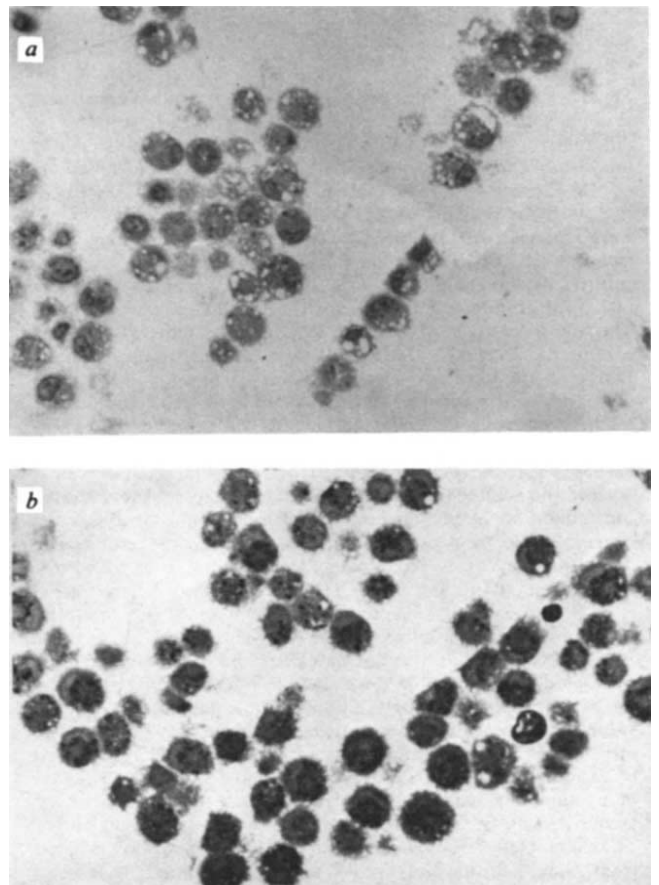


Fig. 2 PMA-treated macrophages (a) show numerous intracytoplasmic vacuoles. In contrast, C5 fragment-treated cells (b) show many cellular processes but less vacuole formation. The cells were fixed in 3% glutaraldehyde and embedded in epoxy resin. Sections (1 μm) were stained with toluidine blue.

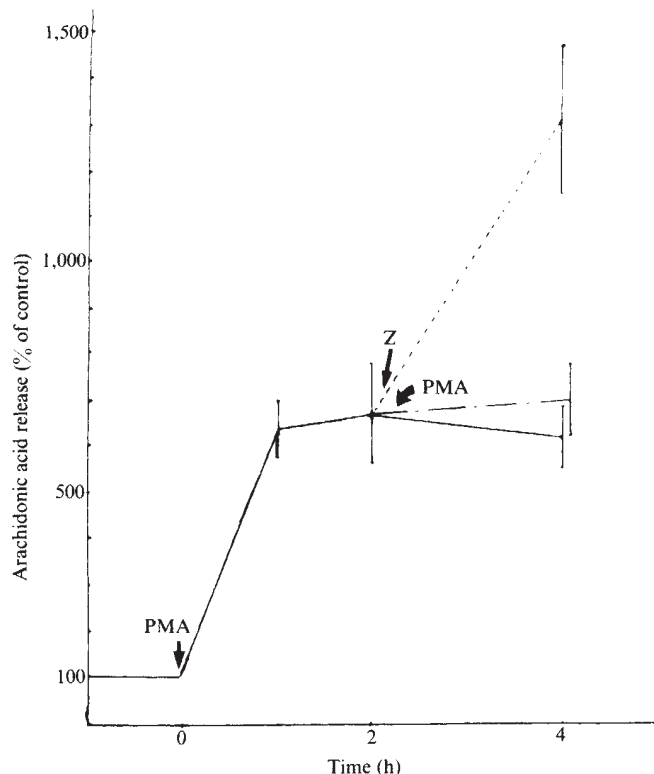


Fig. 3 Time course of arachidonic release by macrophages following various stimuli. At time zero, cells were stimulated with PMA (10^{-6} M). Solid line represents time course of arachidonic acid release following addition of PMA at time zero. Note flattening of the curve after 2 h of incubation. The shape of the curve was not altered despite a new dose of PMA (10^{-6} M) at 2 h (---). However, addition of zymosan (Z, 1 mg ml^{-1}) at this time caused a sharp increase of arachidonic acid release (— — — — —). All values are mean $\pm$ s.e.m., $n = 3$.

addition of zymosan with C5 fragment did not cause much more increase in arachidonic acid production than with zymosan alone. Note also that prostaglandin synthesis following combined zymosan and PMA treatment was only mildly elevated compared with either stimulus alone (Table 1). A possible explanation is saturation of cyclooxygenase, as both PMA and zymosan were used at near maximal dose in this experiment.

Our hypothesis may also be reconciled with the data illustrated in Fig. 3. Repeated stimulation of macrophages with PMA resulted in flattening of the arachidonate releasing curve, suggesting exhaustion of the non-lysosomal phospholipase pool. Addition of zymosan at this time elicited a second sharp increase of arachidonic acid release, indicating the activation of a different pool of phospholipases, namely the lysosomal phospholipases.

This work was supported by NIH grant 1 R23 AI 16147 and a grant from the American Lung Association. Dr John Pike (Upjohn) provided the prostaglandin standards.

Received 23 May 1980; accepted 28 January 1981.

- Humes, J. L. *et al.* *Nature* **269**, 149–151 (1977).
- Hseuh, W., Kuhn, C. III & Needleman, P. *Biochem. J.* **184**, 345–354 (1979).
- Brune, K., Glatt, M., Källin, H. & Peskar, B. A. *Nature* **284**, 261–263 (1978).
- Hseuh, W., Gonzalez-Crussi, F. & Hanneman, E. *Nature* **283**, 80–82 (1980).
- Grimm, W., Seitz, M., Kirchner, H. & Gerns, D. *Cell. Immun.* **40**, 419–426 (1978).
- Hseuh, W. *Am. J. Path.* **96**, 137–147 (1979).
- Bonney, R. J., Naruns, P., Davies, P. & Humes, J. L. *Prostaglandins* **18**, 605–616 (1979).
- Passwell, J. H., Dayer, J. M. & Merler, E. *J. Immun.* **123**, 115–120 (1979).
- Levine, L., Pong, S. S., Hong, S. C. & Tam, S. in *Biochemical Aspects of Prostaglandins and Thromboxanes* (eds Kharasch, N. & Fried, J.) 15–38 (Academic, New York, 1977).
- Flower, R. J. & Blackwell, G. J. *Biochem. Pharmacol.* **25**, 285–291 (1976).
- Vogt, W. in *Advances in Prostaglandin and Thromboxane Research* Vol. 3 (eds Galli, C., Galli, G. & Porcellati, G.) 89–95 (Raven, New York, 1978).
- Kennerly, D. A., Sullivan, T. J., Sylvester, P. & Parker, C. W. *J. exp. Med.* **150**, 1039–1044 (1979).
- Bell, R. L., Kennerly, D. A., Stanford, N. & Majerus, P. W. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3238–3241 (1979).
- Isakson, P. C., Raz, A., Denney, S. E., Wyche, A. & Needleman, P. *Prostaglandins* **14**, 853–871 (1977).

- Bonney, R. J. *et al.* *Biochem. J.* **176**, 433–442 (1978).
- Franson, R. C. & Waite, M. J. *Cell Biol.* **56**, 621–627 (1973).
- McCarthy, K. & Henson, P. M. *J. Immun.* **123**, 2511–2517 (1979).
- Kataoka, K., Ramwell, P. W. & Jessup, S. *Science* **157**, 1187–1189 (1967).
- Bohman, S. O. & Larsson, C. *J. ultrastruct. Res.* **50**, 367 (1975).
- Brune, K., Glatt, M. & Källin, H. *Nature* **274**, 261–263 (1978).
- Ohuchi, K. & Levine, L. *J. biol. Chem.* **253**, 4783–4790 (1978).
- Brinckerhoff, C. E., McMillan, R. M., Fahey, J. V. & Harris, E. D. *Arthritis Rheum.* **22**, 1109–1116 (1979).
- Levine, L. & Hassid, A. *Biochem. biophys. Res. Commun.* **79**, 477–484 (1977).
- Levine, L. & Ohuchi, K. *Nature* **276**, 274–275 (1978).
- Hamberg, M. & Samuelsson, B. *J. biol. Chem.* **241**, 257–263 (1966).
- Goetzl, E. J., Woods, J. M. & Gorman, R. R. *J. clin. Invest.* **59**, 179–183 (1977).
- Desai, U., Krutzer, D. L., Showell, H., Arroyave, C. V. & Ward, P. *Am. J. Path.* **96**, 71–84 (1979).

Antigenic structure of influenza virus haemagglutinin defined by hybridoma antibodies

Walter Gerhard, Jonathan Yewdell & Mark E. Frankel

The Wistar Institute of Anatomy and Biology, Philadelphia, Pennsylvania 19104, USA

Robert Webster

St Jude Children's Research Hospital, Memphis, Tennessee 38101, USA

The recurrence of influenza virus infection in man is attributed primarily to changes occurring in the antigenic structure of the viral surface glycoproteins, especially of the haemagglutinin (HA) molecule^{1–4}. Comparative antigenic analysis⁵ of epidemic influenza virus strains has allowed the description of 'strain-specific' and 'cross-reactive' antigenic determinants^{6–8}. However, the interpretation of these findings remained ambiguous, because the specificity of the applied antisera was insufficiently defined and because the antigenic differences among the HA molecules of various epidemic virus strains resulted presumably from a large number of amino acid substitutions^{9–11}. Thus, in characterizing the antigenic structure of the HA molecule, our approach has been (1) to generate a panel of monoclonal anti-HA hybridoma antibodies, (2) to use some of these antibodies to select mutants of the influenza A/PR/8/34 (PR8) virus expressing antigenically altered HA molecules, and (3) to construct an operational antigenic map of the HA molecule by comparative antigenic analysis of the mutant viruses with the monoclonal antibodies. As we report here, analysis of the 34 mutant viruses selected has enabled us to define four antigenic sites on the HA molecule. Our observation that these sites have undergone antigenic drift to a different extent in nature implies that the mechanisms responsible for antigenic drift act selectively on distinct structures of the HA molecule.

Fifty-eight anti-HA hybridoma antibodies were produced by fusion of lymphocytes from spleen and/or mediastinal lymph nodes of 12 BALB/c mice after intraperitoneal or intranasal inoculation with influenza virus¹². PR8 virus mutants with antigenic alterations (each resulting probably from a single point mutation¹³) in the HA molecule were then obtained by incubating cloned PR8 virus seed with selected hybridoma antibodies and by growing the non-neutralized virus fraction in embryonated hen eggs or in the allantois on shell culture system^{14,15}. Antigenic analysis performed by radioimmunoassay (RIA)¹⁶ on each set of mutant viruses (exemplified in Table 1 with mutants (CV) selected with antibody Sa-1) clearly showed that the non-neutralized viruses are antigenic mutants exhibiting an alteration in the HA molecule which reduced the capacity of the selecting antibody (in present example Sa-1) to bind to the mutant viruses in the RIA. Several antigenically distinct mutant viruses were usually selected by any given hybridoma antibody. For example, antibodies Sa-8 and Sa-9 show that mutants CV2 and CV1 are antigenically unique, whereas antibody Sa-14 shows that CV5 differs from CV6. As no monoclonal antibody

Table 1 Antigenic analysis of the mutant viruses selected in the presence of antibody Sa-1

Hybridoma antibody	Amount of antibody (log 2) binding in RIA to PR8 and its mutants				
	PR8-Mt.S. (parental)	CV1	CV2 (CV9)	CV5	CV6 (CV7, CV8, CV10)
Sa-1 (C)	=6.64	<1	<1	<1	<1
Sa-8	=6.64	6.46 ± 0.24	1	6.26 ± 0.45	6.58 ± 0.18
Sa-14	=6.64	<1	2.48 ± 0.98	<1	6.35 ± 0.02
Sa-9	=6.64	2.09 ± 0.94	6.58 ± 0.31	5.71 ± 0.25	5.78 ± 0.65
36 antibodies	=6.64	6.41 ± 0.24	6.36 ± 0.20	6.40 ± 0.13	6.54 ± 0.12

PR8 mutants were selected by titration of PR8-Mt.S. virus seed in the presence of anti-HA hybridoma ascitic fluid in the allantoic cavity of 10-day-old embryonated hen eggs¹⁴ or in the allantois on shell culture system¹⁵. Virus grown at the titration end point was collected and recloned twice in the absence of antibody in embryonated hen eggs. The cloned viruses were grown up to large quantity, purified and quantitated by HA titration. Viral immunoadsorbents were prepared as follows: 250 µl of HAS (phosphate-buffered saline, 0.04% sodium azide) containing 8×10^4 haemagglutinating units (HAU) of purified virus was mixed with 250 µl of HAS containing 2% NP40, incubated for 30 min at 37 °C and then diluted 1:10 with HAS. These viral NP40 stock solutions could be stored at 4 °C without detectably losing their antibody-binding capacity. The viral immunoadsorbents were prepared by incubating wells of polyvinyl microtitre plates overnight at room temperature with 25-µl aliquots of 1:100 dilutions of the viral NP40 stock solutions (corresponding to 4 HAU of undisrupted virus per well). The plates were incubated for 1 h with HAS containing 1% bovine serum albumin before use in RIA. Supernatants from hybridoma cultures were titrated in the RIA in serial 1:2 dilution steps against the parental viral immunoadsorbent. A standard binding curve was generated in this way for each antibody. One dilution of each antibody (upper part of approximately linear range of the standard curve) was then tested in the RIA against parental and the mutant viral immunoadsorbents. The c.p.m. observed in the parental interaction were defined as 100% (log 2 = 6.64). The antibody-mutant interactions were then expressed as a fraction (log 2) of the standard binding curve of the antibody-parental virus interaction. In each assay a given antibody was tested in parallel on the parental virus (quadruplicate samples) and on the individual mutants (duplicates) of a mutant set. Each assay was performed at least three times. The data in Table 1 are expressed as the mean ± standard deviation of three or four assays. The average standard deviation of antibody-mutant binding values (in the binding range of 4.64–6.64) is 0.37. This value was computed from 35 individual antibody-mutant interactions that were repeated independently of each other over ~1 yr.

was able to discriminate between either CV2 and CV9 or between CV6, CV7, CV8 and CV10, it is possible that the mutant viruses within each of these groups exhibit the same mutation. Lastly, the finding that 36 of the 58 anti-HA antibodies reacted equally well with parental virus and any of the C variants indicates that the antigenic changes induced by individual mutations represent minor modifications in a largely conserved antigenic background.

The observation that individual mutations concomitantly modify certain epitopes¹⁷ (structure on the antigen directly complementary to the combining site of a given antibody) provides an operational basis for the construction of an epitope map^{18,19}. Thus, in the example given above, a relationship between the epitopes delineated by antibodies Sa-1, Sa-8, Sa-9 and Sa-14 is clearly shown by the observation that these epitopes were changed by one or several of the mutations exhibited by the CV mutant viruses. This approach for mapping the antigenic structure of the HA molecule is limited, however, by the number and types of distinct antigenic mutants available for comparative analyses. Thus, the epitopes delineated by the 36 hybridoma antibodies which reacted equally well with the parental virus and the CV mutants could not be mapped on the basis of the CV-mutant virus analysis. We have proceeded, therefore, to select from the same cloned PR8 virus seed additional mutants in the presence of 10 other monoclonal anti-HA antibodies. Each set of mutants was then subjected to an antigenic analysis as exemplified in Table 1 with the CV mutants. A total of 75 PR8 mutants were analysed, of which 34 exhibited a unique antigenic modification. The other 41 mutants could not be further differentiated from a corresponding unique mutant by the present antibody panel and thus probably represent multiple isolates of those unique mutants.

The composite results of the comparative antigenic analyses of the mutants are shown schematically in Fig. 1 where each row represents the reactivity of an anti-HA antibody in RIA with the parental virus (column 1) and the 34 antigenically unique mutants (columns 2–35). Three antibody-binding values are distinguished, black, cross-hatched and white squares representing >50, 25–50 and <25% of the binding to parental virus, respectively. All the hybridoma antibodies, except antibodies Cx-1 to Cx-8, exhibit reduced binding to one or several of the virus mutants. Antibodies and mutant viruses could thus be assembled into four complementary antibody-virus mutant groups designated Sa, Sb, Ca and Cb (the choice for the designations will become clear in the next section). For instance, the mutant viruses CV, KV and PV were assembled into the mutant group Sa while complementary antibodies delineating epitopes that were predominantly modified by antigenic changes exhibited by this mutant group were assembled into a family of antibodies accordingly designated Sa-1 to Sa-15.

The resulting epitope map indicates that each mutant group characterizes an antigenic structure of the HA molecule which can undergo largely independent antigenic alterations. For example, mutations in the Sa mutant group affect the binding capacity of the Sa but not of the Ca antibody family and vice versa. Thus, adopting Urbanski and Margoliash's²⁰ definition of a discrete antigenic site as one to which the binding of specific antibodies is not affected by residue changes at the neighbouring sites, we tentatively designated these structures of the HA molecule as sites Sa, Sb, Ca and Cb. We stress that the reciprocal relationship between antibody families and virus mutant groups is only compatible with the notion that the mapping of antibodies is based on differences in antibody specificity and not on differences in avidity of antibodies directed against the same antigenic site. Also, the very low frequency of mutants selected in the presence of two antibodies directed against different sites further supported the notion of the independent mutability of these sites (unpublished observations).

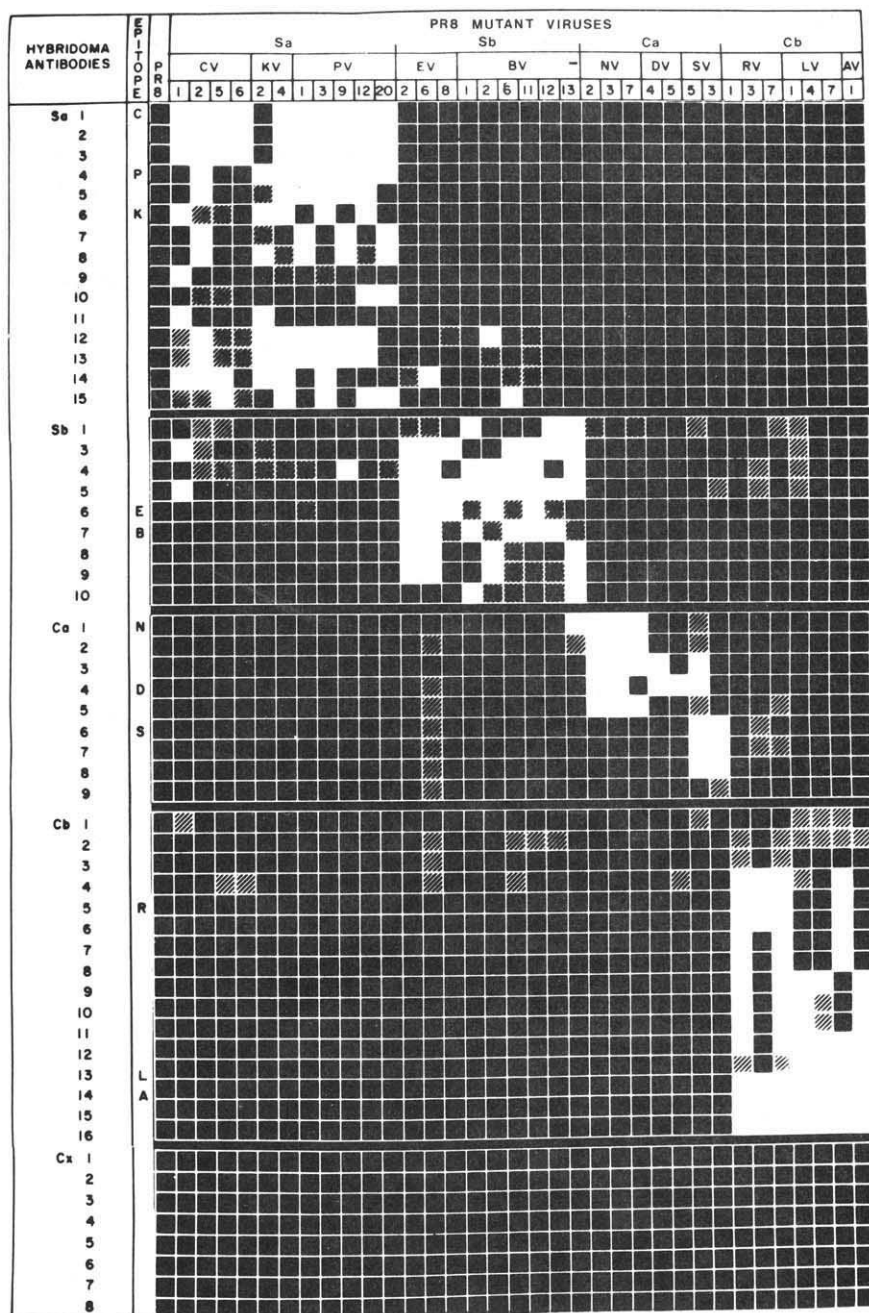
Although most antibodies could be mapped to a single antigenic site, a few exhibited reduced binding to mutants assigned to different sites. In most cases, the observed overlap was between two sites (especially Sa and Sb) and, as discussed below, probably indicates the close physical proximity of these sites. In a few cases, however, linkages among three or four sites were observed. Notably, these linkages were only evident at the level

Table 2 The frequency of mutants in individual sites

Hybridoma antibody	No. of virus samples analysed	Frequency of mutants	
		In individual HA epitopes (log 10, mean ± s.e.m.)	Average per HA site
Sa-1 (C)	4	-5.59 ± 0.09	
Sa-4 (P)	1	-6.49	-6.14 ± 0.41
Sa-6 (K)	2	-6.08 ± 0.60	
Sa-15	3	-6.41 ± 0.33	
Sb-6 (E)	2	-5.32 ± 0.19	-5.48 ± 0.22
Sb-7 (B)	4	-5.63 ± 0.18	
Ca-1 (N)	5	-5.21 ± 0.34	
Ca-4 (D)	5	-4.86 ± 0.21	-5.31 ± 0.50
Ca-6 (S)	1	-5.85	
Cb-13 (L)	4	-6.26 ± 0.19	-5.98 ± 0.40
Cb-5 (R)	2	-5.70 ± 0.67	

Cloned PR8 virus was titrated in the allantois on shell culture system in the absence and presence of an 'over-neutralizing' dose of the indicated anti-HA antibodies¹⁵. Viruses that grew at the titration end point in the presence of anti-HA antibody were shown to be mutants by subsequent antigenic analysis. The mutant frequency (ratio of virus titre in the presence of anti-HA antibody to virus titre in its absence) was determined in several independently cloned PR8 virus samples. Letters in parentheses indicate the viral mutants selected (see Fig. 1).

Fig. 1 Binding of monoclonal antibodies to parental and mutant viral immunoadsorbents in RIA. The hybridomas were produced by fusion of lymphocytes from spleen or mediastinal lymph node cells of 12 BALB/c mice after primary or secondary inoculation with influenza virus by the intravenous, intraperitoneal or intranasal routes¹². The specificity of the hybridoma antibodies for the HA molecule of PR8 was determined by measuring the antibody-binding capacity to recombinant viruses containing reassortments of the genes of PR8 and Hong Kong strains³¹. Antibodies used for mutant selection and the corresponding set of virus mutants are indicated by capital letters. For example, antibody Sa-4(P) was used to select the PV mutants; antibody Sa-6(K) the KV mutants. Black squares represent no detectable decrease in antibody binding to mutants compared with parental virus (binding >5.64, see also Table 1 legend); cross-hatched squares indicate slightly decreased binding (4.64–5.64); white squares indicate binding of <4.64 (in most cases, no detectable binding).



of slight reductions in antibody-binding capacity, which could be due to an unusual sensitivity of these antibodies to alterations of their epitopes as a result of amino acid substitutions in distant antigenic sites.

Caution must be taken in correlating the present antigenic map with the physical structures of the HA molecule, as the epitope map was constructed on the basis of linkages that are operationally but not physically defined. Nonetheless, based on the extent of linkages one could envisage, for example, the antigenic structure of the HA molecule as a large contiguous antigenic area in which the Sa and Sb sites represent neighbouring structures, whereas the Sa and Cb sites correspond to distant structures. Competitive antibody-binding studies²¹ support this general relationship between the operational and topological antigenic map. Further insight into the structural correlates of the present antigenic map could be provided by dissection of the HA molecule into antigenically active protein fragments^{22–25} or determination of the position of the changed amino acid residue in individual mutant HA molecules^{13,26,27}.

Lastly, the antibodies of the Cx family did not discriminate between mutants and parental virus. Final mapping of the corresponding antigenic area will require the selection of mutants with these antibodies. The Cx antibodies may eventually be found to recognize epitopes contained within one of the four defined antigenic sites or alternatively, one or more may define a new antigenic site.

Wilson *et al.*²⁸ have recently resolved the structure of the HA molecule of the influenza virus A/Hong Kong/68 by X-ray crystallography. Amino acid substitutions that had occurred in nature among epidemic virus strains of the Hong Kong subtype and in mutant viruses selected *in vitro* with monoclonal antibodies, clustered into four major areas of the HA molecule²⁹. These areas seemed to represent the structural correlates of four antigenic sites²⁹ which is completely consistent with our findings.

Figure 2 shows the reactivities of our various monoclonal antibodies with epidemic influenza virus strains originally isolated during the period 1931–57 (PR8 was isolated in 1934). The five antibody families cross-react to different degrees with

these heterologous virus strains, suggesting a varying extent of antigenic drift in nature. The Sb antibody family is almost completely nonreactive with virus strains isolated after PR8. Thus, the antigenic site delineated by the Sb antibodies underwent drastic antigenic drift in nature while drift occurred to a slightly lesser extent in the Sa site, in which four cross-reacting antibodies react with heterologous viruses isolated shortly before (SW and WSN, in 1931 and 1933) or shortly after (MEL and BH, in 1935) the isolation of PR8. In contrast, drift occurred

in nature to a much lesser extent in the antigenic sites delineated by the Ca, Cb and Cx antibodies. (Based on these observations we have designated the PR8 'strain-specific' antigenic sites Sa and Sb and the 'cross-reactive' sites Ca and Cb. Note, however, that the Sa and Cb sites contain both 'specific' and 'cross-reactive' epitopes, although at different proportions.)

If antigenic drift in nature results from the selection of spontaneously arising mutant viruses that exhibit a favoured growth potential amid a pre-existing immunity in humans⁴, the mechanisms responsible for antigenic drift clearly act differently on the various antigenic sites defined here. This phenomenon is probably not due to an inherently greater mutability of the antigenic structures corresponding to the Sa and Sb sites as the average frequency of antigenic mutants present in cloned egg-grown PR8 virus is similar in 'specific' and 'cross-reactive' antigenic sites (Table 2). Therefore, the antiviral immune mechanisms seem to exert a stronger selection pressure for mutations in the Sa and Sb sites than the Ca and Cb sites. This could result from a predominant immune response of humans to the Sa and Sb sites and/or greater protective antiviral activities mediated by these antibodies.

Preliminary experiments showed that antibodies directed against the Sa and Sb sites exhibit *in vitro* a >10-fold higher virus-neutralizing potency than antibodies directed against the Cb site, which could explain the more extensive drift in the Sa and Sb sites than in the Cb site. However, very little drift occurred in the Ca site despite the relatively high virus-neutralizing potency of anti-Ca antibodies *in vitro*. Perhaps humans are unable to produce antibodies against the Ca site in sufficient quantities, or anti-Ca antibodies are ineffective *in vivo*.

We have thus used hybridoma antibodies and virus mutants selected *in vitro* to characterize the antigenic topology of the PR8 HA molecule. Webster and Laver¹⁹ have used a similar approach to demonstrate the existence of at least three non-overlapping antigenic sites on the HA molecule of influenza virus A/Mem/71. This approach has so far been used only to investigate the antigenic structure of the influenza virus HA molecule but it is clearly applicable to other viral proteins and to surface proteins of eukaryotic and prokaryotic organisms³⁰. Thus, it should be possible to determine the antigenic structure of a wide variety of biologically relevant antigens, which in turn may be useful for understanding the relationship between specificity and function of immune effector mechanisms.

We thank Maureen Carey, Dale McCreedy and Alex Taylor for help with these experiments. The study was supported in part by grants from the National Institute of Allergy and Infectious Diseases (AI 13989), the National Multiple Sclerosis Society (RG 851C6) and the Medical Scientist Training Program (GM 07170).

Received 10 November 1980; accepted 23 February 1981.

1. Dowdle, W. R., Coleman, M. T. & Gregg, M. B. *Prog. med. Virol.* **17**, 91-135 (1974).
2. Stuart-Harris, C. *Br. med. Bull.* **35**, 3-8 (1979).
3. Kilbourne, E. D. in *The Influenza Viruses and Influenza* (ed. Kilbourne, E. D.) 483-538 (Academic, New York, 1975).
4. Webster, R. G. & Laver, W. G. in *The Influenza Viruses and Influenza* (ed. Kilbourne, E. D.) 270-315 (Academic, New York, 1975).
5. Reichlin, M. *Adv. Immun.* **20**, 71-123 (1975).
6. Schild, G. C. *J. gen. Virol.* **9**, 191-200 (1970).
7. Laver, W. G., Downie, J. C. & Webster, R. G. *Virology* **59**, 230-244 (1974).
8. Virelizier, J. L., Allison, A. C. & Schild, G. C. *J. exp. Med.* **157**, 1571-1578 (1974).
9. Laver, W. G., Air, G. M., Doppeide, T. A. & Ward, C. W. *Nature* **283**, 454-457 (1980).
10. Verhoeven, M. *et al. Nature* **286**, 771-776 (1980).
11. Gething, M.-J., Bye, J., Skehel, J. & Waterfield, M. *Nature* **287**, 302-306 (1980).
12. Gerhard, W., Yewdell, J. W., Frankel, M. E., Lopes, A. D. & Staudt, L. in *Monoclonal Antibodies* (eds Kennett, R. H., McKearn, T. J. & Bechtol, K. B.) 317-333 (Plenum, New York).
13. Laver, W. G. *et al. Virology* **98**, 226-237 (1979); *Proc. natn. Acad. Sci. U.S.A.* **76**, 1425-1429 (1979).
14. Gerhard, W. & Webster, R. G. *J. exp. Med.* **148**, 383-392 (1978).
15. Yewdell, J. W., Webster, R. G. & Gerhard, W. *Nature* **279**, 246-248 (1979).
16. Frankel, M. & Gerhard, W. *Molec. Immun.* **16**, 101-106 (1979).
17. Jerne, N. K. *Annls Immun. Inst. Pasteur, Paris* **1250**, 373-389 (1974).
18. Gerhard, W., Yewdell, J. W. & Frankel, M. E. in *Structure and Variation in Influenza Virus* (eds Laver, W. G. & Air, G. M.) 273-281 (Elsevier, Amsterdam, 1980).
19. Webster, R. G. & Laver, W. G. *Virology* **104**, 139-148 (1980).
20. Urbanski, G. J. & Margoliash, E. *J. Immun.* **118**, 1170-1180 (1977).
21. Lubeck, M. & Gerhard, W. *Virology* (in the press).
22. Eckert, E. A. *J. Virol.* **11**, 183-192 (1973).
23. Brand, C. M. & Skehel, J. J. *Nature* **238**, 145-147 (1972).

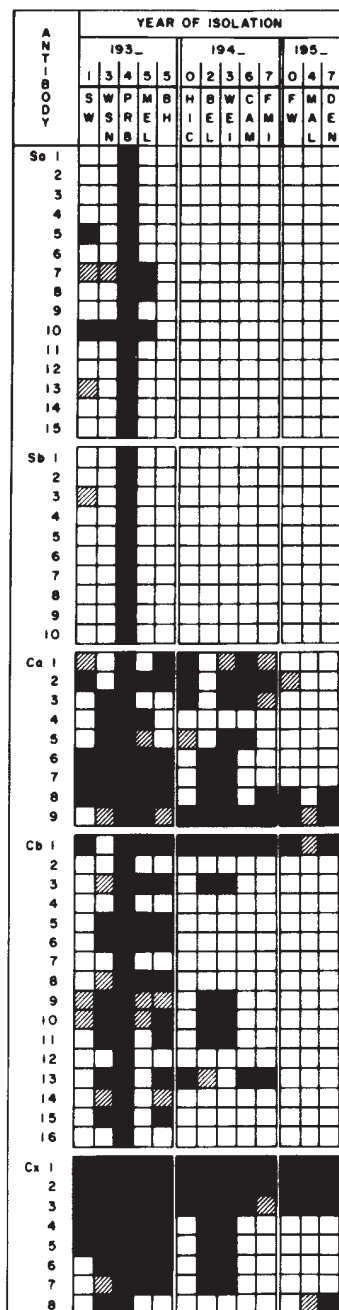


Fig. 2 Binding of hybridoma antibodies in RIA to epidemic influenza virus isolates. The data are similar to those of Fig. 1 except that the white squares represent binding of <2.64, the cross-hatched squares 2.64-4.64, and the black squares >4.64. Wider binding ranges than those in Fig. 1 were chosen so as to differentiate between low (2.64-4.64) and non-detectable (<2.64) cross-reaction. The virus strains used as immunoadsorbents are arranged according to the year of their isolation: A/SW/31(H1N1); A/WSN/33(H1N1); A/PR/8/34(H1N1); A/Melbourne/35(H1N1); A/BH/35(H1N1); A/Hickox/40(H1N1); A/Bellamy/42(H1N1); A/Weiss/45(H1N1); A/CAM/46(H1N1); A/FM/1/47(H1N1); A/FW/1/50(H1N1); A/Malaya/302/54(H1N1); A/Denver/1/57(H1N1).

24. Jackson, D. C., Dopheide, T. A., Russell, R. J., White, D. O. & Ward, C. W. *Virology* **93**, 458–465 (1979).
25. Jackson, D. C., Brown, L. A., White, D. O., Dopheide, T. A. & Ward, C. W. *J. Immun.* **123**, 2610–2617 (1979).
26. Laver, W. G. *et al.* in *Structure and Variation in Influenza Virus* (eds Laver, W. G. & Air, G. M.) 295–306 (Elsevier, Amsterdam, 1980).
27. Moss, B. A., Underwood, P. A., Bender, V. J. & Whittaker, R. G. in *Structure and Variation of Influenza Virus* (eds Laver, W. G. & Air, G. M.) 329–338 (Elsevier, Amsterdam, 1980).
28. Wilson, I. A., Skehel, J. J. & Wiley, D. C. *Nature* **289**, 366–373 (1981).
29. Wilson, D. C., Wilson, I. A. & Skehel, J. J. *Nature* **289**, 373–378 (1981).
30. Holtkamp, B., Cramer, M., Lemke, H. & Rajewsky, K. *Nature* **289**, 66–68 (1981).
31. Yewdell, J. W. & Gerhard, W. A. *Rev. Microbiol.* (in the press).

Short-lived cytoplasmic regulators of gene expression in cell cybrids

C. Ronald Kahn*, Roger Bertolotti, Maryvonne Ninio & Mary C. Weiss

Centre de Genetique Moleculaire du CNRS,
91190 Gif-sur-Yvette, France

Somatic cell hybridization is a valuable tool for investigating the control of gene expression in eukaryotic cells^{1–4}. Studies of hybrid cells, heterokaryons, reconstructed cells and cybrids (cytoplasmic hybrids) have suggested that cytoplasmic factors may be involved in this regulatory process. Unfortunately, studies of this kind usually require that hybrid or modified cells be maintained for some time in a selective environment during which chromosomal losses or other changes may modify the genetic functions of the cells and thus vitiate conclusions about the mechanism of gene regulation. We report here the preparation of cybrids between enucleated mouse fibroblasts (Cl-1-D) and differentiated rat hepatoma cells (Fao) and the use of a combination of histological techniques to identify these modified cells early after fusion without the use of selective media. We found that albumin production in most cybrids was suppressed (extinguished) at 12–20 h after fusion but was restored by 48 h. These results suggest that there is a cytoplasmic factor in the fibroblast which exerts negative control over expression of the albumin gene, but which in the absence of the fibroblast nucleus, is not renewed and therefore short-lived.

The loss of gene functions characteristic of differentiation is commonly observed after cell fusion and is termed extinction^{1,2,4–6}. For example, most hybrids between fibroblasts and hepatoma cells do not express differentiated liver functions such as albumin synthesis, tyrosine aminotransferase and aldolase. The phenomenon, well studied in hybrids, has also been observed in heterokaryons (refs 7–10 and M. Ninio and M. Weiss, in preparation), and it has been suggested that differentiated gene functions are controlled by diffusible substances whose effects are negative or suppressive^{1–3}. Evidence that the substances concerned may be cytoplasmic has been provided by experiments¹¹ in which cybrids obtained by fusing Friend erythroleukaemia cells with cytoplasts of neuroblastoma cells or fibroblasts failed to produce haemoglobin in circumstances in which erythroleukaemic cells would have done so. However, most attempts to transfer extinction or other properties under genetic control by cybridization have failed (refs 12–14 and C. R. Kahn, T. V. Gopalakrishnan and M. C. Weiss, in preparation). Such negative experiments, however, do not rule out the possibility of cytoplasmic regulators of gene expression, because the cybrids usually cannot be examined until several days to weeks after fusion when grown in selective conditions. Here we used a combination of histological techniques to search for cytoplasmic regulatory factors early after cybrid formation.

The cytoplasm of the mouse fibroblast Cl-1-D was labelled by allowing the cells to phagocytose polystyrene beads (0.5–1.0 µm), after which enucleation was performed using the method of Gopalakrishnan and Thompson¹⁵. Cl-1-D cells are somewhat resistant to cytochalasin-induced enucleation, but with this method between 50 and 70% of the cells can be enucleated with high yield. After a short period of recovery, the cytoplasts, still attached to the plastic disks used for enucleation, were fused with the differentiated rat hepatoma, Fao, using polyethylene glycol (PEG). At selected time intervals, the cells were fixed and stained by an indirect immunofluorescent technique for albumin⁹ followed by nuclear staining with the fluorochrome dye Hoescht 33258 (ref. 16). The cells were then examined using phase contrast and epifluorescent illumination.

This combination of histological techniques allowed the identification of both parent cell types, heterokaryons, hybrids and cybrids. The fibroblast cytoplasm was labelled with beads and the hepatoma was not; the rat (hepatoma) nuclei stained homogeneously with Hoescht 33258, whereas mouse (fibroblast) nuclei possessed densely staining heterochromatin or chromocentres which fluoresced as bright spots (Fig. 1, top and middle rows respectively). Heterokaryons possessed cytoplasmic beads and both rat and mouse nuclei; hybrids, cytoplasmic beads and a large 'mixed-type' of nucleus; and cybrids, cytoplasmic beads and only a rat nucleus (or nuclei) (Fig. 1, bottom row). Each of these cell types was evaluated for albumin production by immunofluorescent staining (Fig. 1c, f, i).

At 12–20 h after fusion, cybrids between the enucleated fibroblast Cl-1-D and the rat hepatoma Fao showed a marked extinction of albumin production based on immunofluorescent staining (Fig. 1i). In four separate experiments, 59–72% of all mononuclear (1N) cybrids examined were negative for albumin, compared with <10% negative in normal cultures of Fao (Table 1; Fig. 2). This is comparable with the frequency of extinction of albumin production observed in the same experiments in heterokaryons between intact Cl-1-D and Fao at this time point (data not shown). The extinction was not due to a toxic effect of the PEG or to the fusion itself because intact Fao cells on the same disk or Fao homokaryons (Fao cells fused with other Fao) were between 89 and 91% positive for albumin by this technique. Furthermore, the extinction was not due to quenching of fluorescence by the polystyrene beads, because Fao cells which had been allowed to phagocytose beads directly showed comparable levels of albumin by immunofluorescence (data not shown).

Extinction was a rapid but transient phenomenon (Fig. 2, Table 1); within 6 h after fusion, 18–35% of all cybrids already demonstrated extinction, and extinction reached its maximum by 12–20 h (Table 1). After 24 h, a decline in the level of extinction was apparent, and by 45–48 h cybrids were 90% positive for albumin, a level comparable with that in the non-fused Fao.

Although we cannot accurately quantitate the ratio of cytoplasm contributed by the enucleated fibroblasts and the intact hepatoma, some attempt was made to evaluate the concentration dependence of the phenomenon by comparing the frequency of extinction in cybrids containing one nucleus with those containing more than one nucleus (but only rat hepatoma nuclei). In each experiment, and at every time point, the frequency of extinction was greater in the one-nucleus cybrids (1N), than in cybrids containing more than one nucleus (>1N) (and thus more than one rat hepatoma cell) (Table 1, Fig. 2b). In general, extinction seemed to be more frequent in those cybrids containing the largest numbers of beads, although this proved impossible to quantitate because cytoplasts varied considerably in the extent of labelling and those more heavily labelled with beads were preferentially lost during the enucleation procedure. However, these experiments strongly suggest that the cytoplasm of the mouse fibroblast contains a diffusible regulatory molecule which is short-lived and can influence the expression of albumin production in a time- and concentration-dependent fashion.

* Permanent address: Section on Cellular and Molecular Physiology, Building 10, Room 8S-243, National Institutes of Health, Bethesda, Maryland 20205, USA.

Despite intensive study, the molecular basis for the regulation of gene expression during the process of cellular differentiation remains poorly understood. It has been assumed that there may be both positive and negative regulators of gene expression in eukaryotic cells analogous to those observed in bacteria¹⁷. The existence of cytoplasmic molecules which may serve as regulators of gene expression has been clearly demonstrated by nuclear transplantation in protozoa and lower vertebrate species^{18,19}.

More recently, evidence has also been provided for the existence of similar positive and negative cytoplasmic regulatory factors in mammalian cells. Analysis of differentiated functions in somatic cell hybrids¹⁻⁶ and heterokaryons (refs 7-10 and M. Ninio and M. Weiss, in preparation) has suggested that the expression of genes in mammalian cells may be regulated by diffusible substances whose final effect may be either negative (extinction) or positive (activation). Evidence that these factors exist in the cytoplasm has been suggested by two types of experiment. Lipisch *et al.*²⁰ observed activation of the gene for tyrosine aminotransferase in reconstructed cells prepared from

mouse fibroblast nuclei and rat hepatoma cytoplasts, and Gopalakrishnan *et al.*^{11,21} have observed both extinction and activation of differentiated functions in cybrids prepared from mouse erythroleukaemia cells and cytoplasts derived from a variety of cells. With these exceptions, however, most attempts to transfer cytoplasmic regulatory factors have yielded negative results¹²⁻¹⁴. For example, we found that 6 weeks after fusion, cell hybrids between hepatoma and fibroblasts almost invariably showed extinction of differentiated liver functions (such as albumin and α -fetoprotein production, aldolase activity), whereas cybrids prepared using only the fibroblast cytoplasm did not, although chloramphenicol resistance (a fibroblast mitochondrial marker) was transferred, indicating that fusion had indeed taken place (C. R. Kahn, T. V. Gopalakrishnan and M. C. Weiss, in preparation).

Our data suggest that the fibroblast cytoplasm does indeed contain a molecule(s) capable of producing extinction of albumin production with very high efficiency. The reason this potential regulator of gene expression was not previously observed is that, in the absence of a nucleus coding for new synthesis, the

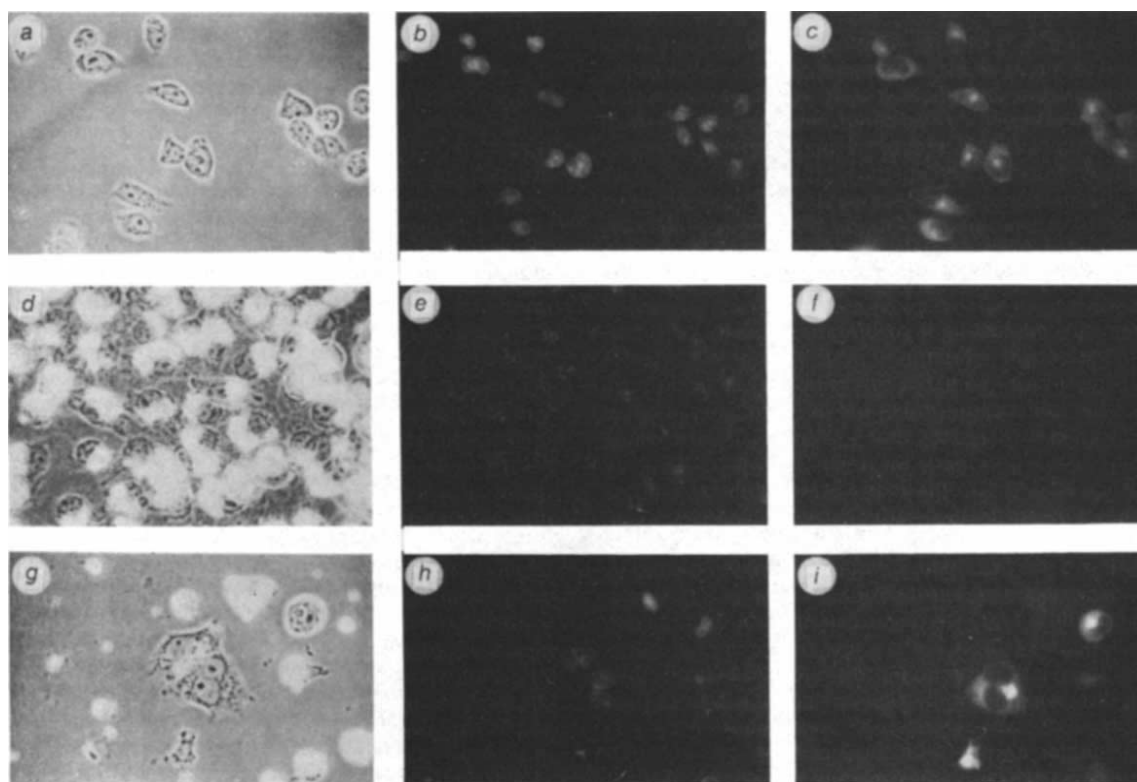


Fig. 1 Morphological and immunological characterization of the parental cells and cybrids. CI-1-D (middle row) is a clonal line of fibroblastic mouse cells, sometimes referred to as LM(TK)⁻²⁸. Fao (top row) is a fourth-generation clonal line of well differentiated rat hepatoma cells derived from the hepatoma line H4IIEC3 which is resistant to 8-azaguanine and ouabain²⁹. Both cell lines were cultivated in modified Ham's F12 medium³⁰ supplemented with 5% fetal calf serum in a humidified atmosphere, 7% CO₂. To prepare cybrids (bottom row), CI-1-D cells were first allowed to phagocytose polystyrene beads (Dow) by adding a sterile suspension to culture medium and incubating the cells overnight. Beads of 0.8 μ m were of optimum size for easy detection without gross distortion of morphology *d*, although 0.5 and 1.0 μ m beads were also usable. The cells were then detached with trypsin and EDTA²⁹, resuspended in phosphate-buffered saline (PBS), and 1 ml of cell suspension ($2-3 \times 10^6$ cells ml⁻¹) was allowed to attach to 45-mm plastic disks precoated with concanavalin A (0.15 mg) and enucleated as described by Gopalakrishnan and Thompson¹⁵ using medium containing 10 μ g ml⁻¹ of cytochalasin B. A 50-70% enucleation could be achieved by centrifugation for 40 min at 16,000g at 37 °C with a ~40-50% yield. After a recovery period of 30-45 min in normal growth medium, the medium was aspirated and 1 ml of a suspension of $2.5-3.5 \times 10^5$ Fao cells was carefully layered onto the plastic disk. After 2-2.5 h at 37 °C in the incubator, the Fao cells were well attached and fusion was carried out by treating for 1 min with a solution of 50% PEG (1000 Merck, Ultrapure) in serum-free medium. The cells were washed four times with serum-free media, then returned to the incubator in 60-mm Petri dishes containing 5 ml of growth medium. At appropriate intervals, the coverslips were removed and prepared for histological examination as described by Ninio *et al.* (in preparation). The cells were fixed in 3% formaldehyde for 1 min, which was diluted 1:1 with methanol, then replaced completely by pure methanol for 30 min. The cells were then incubated with a rabbit anti-rat albumin serum (dilution 1:100) for 30 min at 37 °C, washed with PBS for 5 min, then stained with a fluorescein-labelled sheep anti-rabbit IgG (Institut Pasteur, dilution 1:200) for 30 min at 37 °C. After an additional wash with PBS, the cells were stained with the fluorochrome Hoechst 33258 (0.5 μ g ml⁻¹)¹⁶ and rinsed again with PBS. Glass coverslips were mounted over the cells with buffered glycerol and sealed with Histolaque. The slides could then be stored at 4 °C or examined directly in phase contrast and epifluorescent illumination using a Zeiss UV microscope equipped with a HBO 200 W mercury lamp. Photographs were taken with either a $\times 40$ or $\times 100$ oil immersion objective, using Ilford HP-5 or Ektachrome film. This figure shows the appearance in phase contrast and fluorescence microscopy of the two parent cell types and the cybrids: *a, d, g*, phase contrast; *b, e, h*, Hoechst 33258; *c, f, i*, albumin. Hoechst 33258 stains the nuclei of the rat hepatoma homogeneously (*b*), whereas the mouse fibroblast nuclei have a dotted appearance due to more intense staining of the chromocentres (*e*). The bottom row shows a typical cybrid next to a binucleated Fao cell or homokaryon. Both cells have only rat type nuclei (*h*), but the cybrid also shows cytoplasmic beads as a result of fusion with a CI-1-D cytoplast. Note that albumin (*i*) is extinguished in the cybrid but not in the adjacent Fao cell.

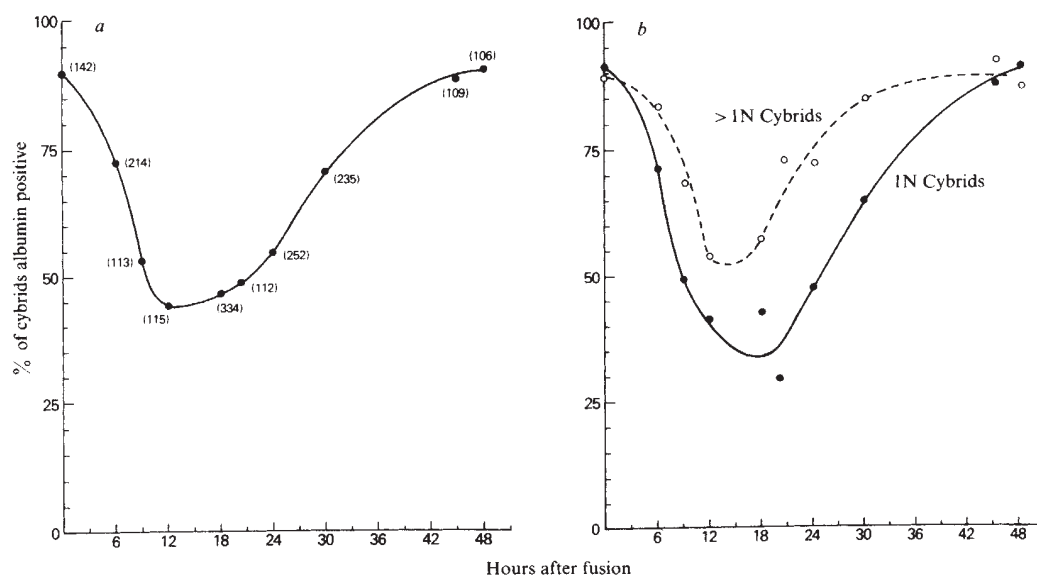


Fig. 2 Time course of extinction of albumin production in cybrids between Cl-1-D fibroblasts and the rat hepatoma Fao. *a*, The experiments in Table 1 have been summarized by totalling all cybrids examined at each time point for the four experiments and calculating the per cent albumin-positive cells by immunofluorescence. Numbers in parentheses indicate the number of cybrids examined. *b*, The data in *a* were subdivided to show the mononuclear cybrids (1N, solid line) versus the cybrids with more than one rat nucleus (>1N, broken line). Note that at every time point extinction is more frequent in the 1N cybrids. We believe that immunofluorescent staining for albumin will be positive until the cellular content of albumin is reduced by >75%.

factor has a relatively short half life. Thus, albumin production was inhibited or markedly repressed in the majority of cybrids by 12–20 h after fusion, but by 48 h extinction was no longer observed, presumably because the factor had been metabolized or diluted by cell division to a level no longer sufficient for activity. This is supported by data obtained in a parallel series of experiments in which heterokaryons between Cl-1-D and Fao cells showed extinction with a similar rate of onset, but which persisted for the 10 days of the study (M. Ninio and M. C. Weiss, in preparation). Although quantitation is difficult, both studies showed a dose-dependent effect; cybrids and heterokaryons with more than one hepatoma nucleus showed less extinction than those with only a single hepatoma nucleus.

Neither the chemical nature of the factor(s) nor the location of regulation is apparent. Previous studies with hybrids and mutant cells have suggested that regulation of gene expression occurs at a pretranslational level and is specific for the differentiated functions of the cell^{1–6}. Studies in prokaryotes have indicated that both proteins and small molecules, such as cyclic AMP, may act as regulatory factors¹⁷. The time course of extinction observed in these experiments is more consistent with the half life

of a protein than a cyclic nucleotide. However, it would not be surprising if the time course of effect varied extensively depending on the half life of the regulator in its new environment, as well as its mode of action. Activation of the tyrosine amino-transferase gene in the reconstructed cells formed from fibroblast nuclei and hepatoma cytoplasm lasted 6–8 weeks²⁰, and permanent activation of the gene for phenylalanine hydroxylase seems to occur in cybrids formed from erythroleukaemia cells and hepatoma cytoplasts²¹. However, using an experimental design similar to that reported here, except that cybrids were formed between enucleated hepatoma cells and fibroblasts, we have been unable to obtain evidence for any positive regulators of albumin gene expression.

Clearly, some cytoplasmic regulators of gene expression exist. The transient effect observed in the absence of the nucleus supports the concept that, in general, the epigenotype is stable¹. Thus, somatic-cell hybrids re-express parental differentiated functions following the loss of chromosomes from the extinguishing parent^{22–25}. It is also evident, however, that, in the eukaryotic cell, there are multiple potential mechanisms of gene regulation. Direct studies of cell extracts using chemical

Table 1 Immunofluorescent quantitation of albumin production in cybrids between Cl-1-D and Fao cells

Expt	Time (h)	No. of 1N cybrids examined		% Of 1N cybrids albumin positive	No. of >1N cybrids examined		% Of >1N cybrids albumin positive	No. of Total cybrids examined		% Of all cybrids albumin positive
		Alb ⁺	Alb ⁻		Alb ⁺	Alb ⁻		Alb ⁺	Alb ⁻	
1	6	64	38	62.3	9	2	81.8	73	40	64.6
	18	38	55	40.9	10	3	76.9	48	58	45.3
	30	58	19	75.3	29	2	93.5	87	19	82.1
2	18	20	43	31.7	18	26	40.9	38	69	35.5
	24	28	34	53.8	24	15	61.5	52	49	51.5
3	9	45	46	49.5	15	7	68.2	60	53	53.1
	20	17	43	28.3	38	14	73.0	55	57	49.1
	45	84	12	87.5	12	1	92.3	96	13	88.1
4	6	67	15	81.7	16	3	84.2	83	18	82.2
	12	31	46	40.3	15	13	53.6	46	59	43.8
	18	53	48	52.0	16	4	75.0	69	52	62.7
	24	53	53	50.0	22	3	88.0	75	56	67.6
	30	57	43	51.8	22	7	75.9	79	50	66.4
	48	76	8	91.6	19	3	86.4	95	11	89.6

fractionation and microinjection²⁶, as well as further characterization of the DNA and specific genes²⁷ from expressing and non-expressing cell lines, should elucidate these problems.

Received 17 October 1980; accepted 15 January 1981.

1. Ephrussi, B. *Hybridization of Somatic Cells* (Princeton University Press, New Jersey, 1972).
2. Weiss, M. C. *Proc. 5th Int. Congr. hum. Genet.*, Mexico City, 284–292 (Excerpta Medica, Amsterdam, 1976).
3. Ringertz, N. R. & Savage, R. E. *Cell Hybrids* (Academic, New York, 1976).
4. Davidson, R. L. A. *Rev. Genet.* **8**, 195–218 (1974).
5. Weiss, M. C., Sparkes, R. & Bertolotti, R. *Somat. Cell Genet.* **1**, 27–40 (1975).
6. Deisseroth, A. *et al. Somat. Cell Genet.* **2**, 373–384 (1976).
7. Gordon, S. & Cohn, Z. J. *exp. Med.* **131**, 981–1003 (1970).
8. Thompson, E. B. & Gelehrter, T. D. *Proc. natn. Acad. Sci. U.S.A.* **68**, 2589–2593 (1971).
9. Harris, H. *Cell Fusion: The Dunham Lectures* (Clarendon, Oxford, 1970).
10. Zeuthen, J. S., Stenman, S., Fabricius, H. A. & Nilsson, K. *Cell Differentiation* **4**, 369–386 (1976).
11. Gopalakrishnan, T. V., Thompson, E. B. & Anderson, W. F. *Proc. natn. Acad. Sci. U.S.A.* **74**, 1642–1646 (1977).
12. Howell, A. N. & Sager, R. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2358–2362 (1978).
13. McBurney, M. W. & Strutt, B. *Expl Cell Res.* **124**, 171–180 (1979).
14. Wright, W. E. & Hayflick, L. *Expl Cell Res.* **96**, 113–121 (1975).
15. Gopalakrishnan, T. V. & Thompson, E. B. *Expl Cell Res.* **96**, 435–439 (1975).
16. Wright, W. E. & Gros, F. *Expl Cell Res.* **111**, 451–454 (1978).
17. Goldberger, R. F., Deeley, R. G. & Mullinix, K. P. *Adv. Genet.* **18**, 1–68 (1976).
18. Danielli, J. F. *Harvey Lect.* **58**, 217–232 (1963).
19. DeRobertis, E. M. & Gurdon, J. B. *Proc. natn. Acad. Sci. U.S.A.* **74**, 2470–2474 (1977).
20. Lipisch, L. A., Kates, J. R. & Lucas, J. J. *Nature* **281**, 74–76 (1979).
21. Gopalakrishnan, T. V. & Anderson, W. F. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3932–3936 (1979).
22. Bertolotti, R. *Somat. Cell Genet.* **6**, 579–602 (1977).
23. Klebe, R. J., Chen, T. R. & Ruddle, F. H. *Proc. natn. Acad. Sci. U.S.A.* **66**, 1220–1227 (1970).
24. Weiss, M. C. & Chaplain, M. *Proc. natn. Acad. Sci. U.S.A.* **68**, 3026–3030 (1971).
25. Croce, C. M., Litwalk, G. & Koprowski, H. *Proc. natn. Acad. Sci. U.S.A.* **70**, 1268–1272 (1973).
26. Diacumakos, F. G. *Meth. Cell Biol.* **7**, 288–312 (1973).
27. Sperling, L. & Weiss, M. C. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3412–3416 (1970).
28. Dubbs, D. R. & Kit, S. *Expl Cell Res.* **33**, 19–28 (1964).
29. Deschatrette, J., Moore, E. E., Dubois, M., Cassio, D. & Weiss, M. C. *Somat. Cell Genet.* **5**, 697–718 (1979).
30. Coon, H. G. & Weiss, M. C. *Proc. natn. Acad. Sci. U.S.A.* **62**, 85–859 (1969).

Polyoma DNA sequences involved in control of viral gene expression in murine embryonal carcinoma cells

Michaël Katinka*, Marc Vasseur†, Nicole Montreau†, Moshe Yaniv* & Daniel Blangy†

* Département de Biologie moléculaire, Institut Pasteur, 25 rue du Dr Roux, 75015 Paris, France

† Laboratoire des Virus oncogènes, Université Pierre et Marie Curie, Institut de Recherches Scientifiques sur le Cancer, BP no 8, 94800 Villejuif, France

Polyoma virus can develop lytically in differentiated mouse cells, but is unable to express either early or late functions in undifferentiated early embryonic cells¹ or in embryonal carcinoma (EC) cells of the mouse. However, the EC cells become susceptible during differentiation and then behave as do mouse somatic cells². We have recently described the isolation of a new class of polyoma virus mutant (Py EC) able to develop normally in an EC cell line (PCC4-Aza) that is capable of trigeminal differentiation *in vivo*³ but exhibits a very limited differentiation *in vitro*⁴. The mutants do not grow in another EC cell line (F9) that is 'blocked'—'nullipotent' *in vivo*, but able to differentiate *in vitro* when grown in the presence of chemical inducers. Compared with their parental strain, all Py EC genomes analysed so far have an extra 20–50 bases in their *HpaII*-3 restriction fragment⁵. We now describe polyoma mutants that will grow in EC F9 cells, and the sequence changes in two of them. The region involved in the mutations can be folded into a putative tRNA-like secondary structure which may be involved in the regulation of viral early transcription with respect to the state of differentiation of the host cell.

Mutants of polyoma virus able to develop on F9 EC cells were obtained using two different procedures: F9 cells were infected (1) with an heterogeneous and highly defective polyoma stock, obtained by viral propagation at high multiplicity of infection (MOI) on permissive mouse cells, as already described for the

selection of Py EC on PCC4-Aza, or (2) with purified wild-type polyoma A2, at a MOI of 5,000 plaque-forming units (PFU) per cell. Following their infection, F9 cells were cultured as usual and scored periodically by immunofluorescence for polyoma T (tumour) and V (late structural proteins) antigen expression. In both procedures, and as reported for the PCC4-Aza cells, viral proteins were detected in 5–20% of the cells 3–4 weeks after infection. The virus produced was recovered and plaque purified three times on secondary mouse embryo cells. (We propose calling the mutants Py EC, followed by the name of the cells on which they were selected and the isolate number, for example, Py EC PCC4-97, Py EC F9-5.) Transforming ability on rat FR3T3 cells⁶, efficiency of tumour induction in newborn hamsters and burst size on various permissive mouse cell lines were checked for independent Py EC F9 isolates. As observed in the case of Py EC PCC4, no significant difference between Py EC F9 mutants and wild-type virus could be found.

Endonuclease restriction analysis of five independent Py EC F9 mutants revealed no common detectable size modifications of the *HpaII*-3 fragment (compare the EC PCC4 mutants of ref. 9) nor any specific rearrangement associated with either of the selection procedures used.

F9 EC cells infected with 10 PFU of any of the Py EC F9 mutants tested, exhibit, 42 h after infection, more than 60% of T-antigen positive cells, as measured by immunofluorescence⁸. In the same conditions of infection, never more than 5% PCC4-Aza cells were shown to express T or V antigen (results not shown).

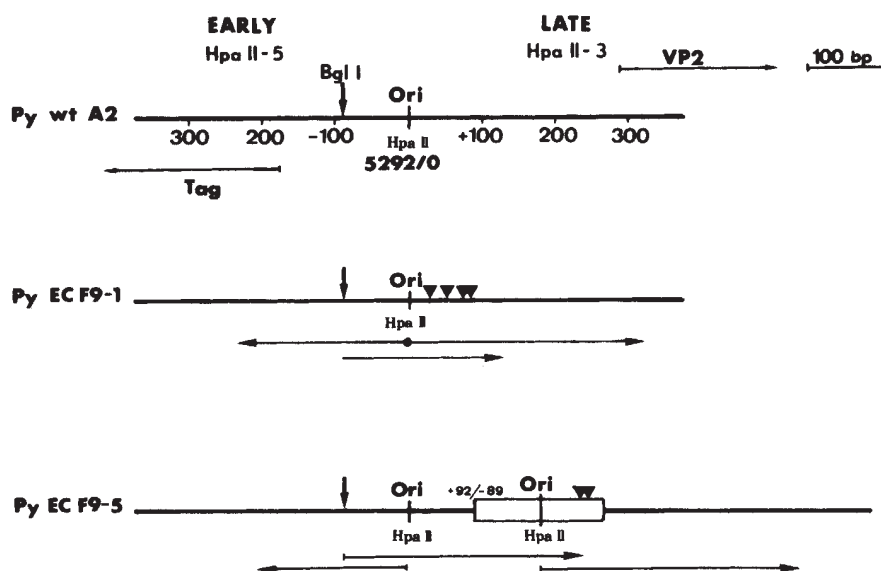
The efficiency of F9 infection by the Py EC F9 mutant can be compared with that of PCC4-Aza infection by Py EC PCC4 mutants⁹: in the latter case, at 37 °C, never more than 40% T- or V-antigen positive cells were observed (after infection at 100 PFU per cell) whereas F9 cells infected at 5–10 PFU with Py EC F9 mutants allow viral expression in 60–80% of the cells. All these biological properties suggest that restriction of polyoma expression in F9 and PCC4 may involve different factors.

The cloned viral genomes inserted into the *Bam*HI site of PBR322 were tested for their infectivity by transfecting secondary mouse embryo cells with viral DNA obtained after cleavage with *Bam*HI and ligation. The virus burst obtained was then used to infect both EC F9 and EC PCC4 cells at a MOI of 10–20 PFU per cell. When monitored for the appearance of T and V antigens by immunofluorescence, 60–80% of the EC F9 cells but only 3–5% of the EC PCC4 cells were positive. DNA extracted from two mutant infectious viruses revealed the expected restriction pattern, without apparent defective genomes.

About 500 nucleotides around the origin of DNA replication were sequenced in Py EC F9-1 and about 700 in Py EC F9-5, as described in the legends to Figs 1 and 2. Only slight differences from the wild-type polyoma sequence were found. Py EC F9-1 has three insertions: a thymine between positions 85 and 86, another thymine after position 109 and an adenine between positions 119 and 120. One transition of a thymine into a cytosine at position 63 is also observed. (The numbering system used here is based on the PyA2 sequence of Arrand *et al.*¹⁰ where residues in the late region are counted from the middle of the *HpaII*-3/5 taken as zero.) In Py EC F9-5, there is a duplication of a 179-base pair region covering the origin of DNA replication but only the right-hand duplicated sequence reveals the existence of two supplementary nucleotides, a thymine inserted between positions 108 and 109 and an adenine between nucleotides 120 and 121. We do not believe that the duplication in Py EC F9-5 has any role in the properties of the mutant because other polyoma mutants with a duplication of the DNA origin of replication¹¹ do not give rise to productive infection on either EC F9 or EC PCC4 cells. The rest of the nucleotide sequence is identical to the one we have determined previously for the A2 wild-type polyoma¹², and differs only slightly from that published by Arrand *et al.*¹⁰. To confirm that the host-range capacities were due solely to the point mutations detected, and that no other modification elsewhere in the polyoma genome

Fig. 1 Genomic organization around the origin of DNA replication in wild-type and mutant polyoma; the DNA sequencing strategy. The origin of DNA replication is commonly given as the *HpaII*-3/5 junction. The beginning of translation of the early T antigens at -173 base pairs and the late VP2 at +291 base pairs, as well as its direction are shown by the corresponding arrows. Underneath are given the parallel regions of the two mutants analysed in the present study. The nucleotide sequence determined is presented by the underlying arrows. For both mutants, the *HpaII*-3, *HpaII*-5 and *BglII*-3 fragments were purified on polyacrylamide gels after digestion of the corresponding pBR322-Py plasmids, and labelled at their 5' end with ^{32}P using the exchange reaction with T_4 polynucleotide kinase as described by Berkner and Folk²⁰.

After secondary restriction with *Bam*HI, *Alu*I and *Bam*HI, respectively, the labelled and purified fragments were submitted to the partial chemical degradation of Maxam and Gilbert²¹ and the subsequent products were fractionated on 6 or 20% polyacrylamide-urea 0.35-mm thin gels²². Autoradiography was done either with Kodak NS-2T film at -20 °C or with an intensifying screen (Cronex Lightning-Plus, DuPont) using a preflashed Kodak X-OMAT R film at -70 °C (ref. 23). Point mutations observed are shown (▼). The duplication in polyoma EC F9-5 is designated by the box.



was involved, we constructed a hybrid DNA molecule in which the mutant *BglII*-*Bam*HI 750-base pair (930 base pairs in Py EC F9-5) fragment covering the mutated region was ligated to the complementing 4.5-kilobase region of wild-type polyoma, and vice versa. The hybrid DNA was used to transfect secondary mouse embryo cells and the virus burst was utilized to

infect EC F9 cells. Only the hybrids containing the 750-base pair (or 930-base pair) fragment from the mutants gave productive infection as assayed by the appearance of the T and V antigens (about 60–80% positive cells compared with 0.1% in the opposite construction). Segal *et al.*¹³ have detected a low amount of unspliced SV40 DNA in F9 cells after infection with

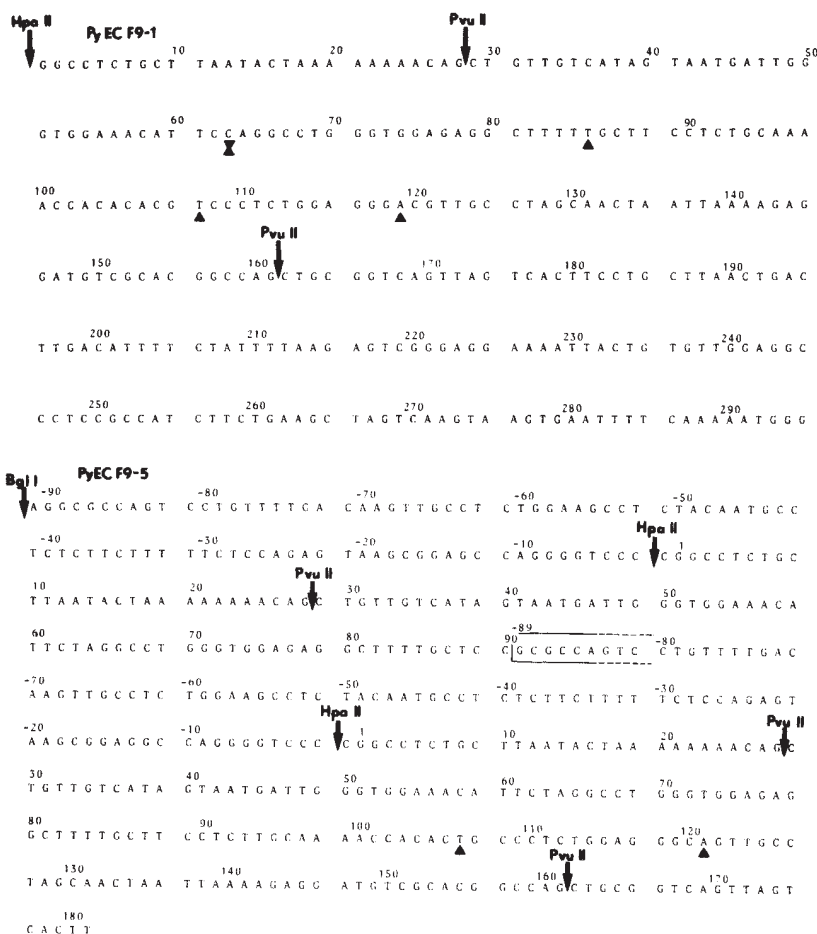


Fig. 2 Nucleotide sequence of Py EC F9-1 and of Py EC F9-5. The insertions (▲) and transitions (■) in the mutant sequences are shown. The restriction sites for *BglII*, *HpaII* and *PvuII* are given as landmarks. The numeration of the nucleotides follows that of Arrand *et al.*¹⁰.

this virus. The fact that the polyoma mutants are not located in the region concerned with the splicing of polyoma early RNA tends to exclude a possible block at this step during infection with wild-type polyoma.

The changes occurring in the Py EC PCC4 and Py EC F9 mutants show that both groups have genomes altered in the sequence of the *HpaII*-3 fragment, near the origin of DNA replication on the late side of the genome, in a noncoding region (Fig. 4). Py EC PCC4 mutants reveal a genomic rearrangement in which a deletion (from nucleotide 46 or 77 to nucleotide 107) is replaced by the duplication of a downstream region (nucleotide 138 or 157 to 220; see Fig. 4 and ref. 12), whereas the Py EC F9 mutants exhibit in the same region a series of point mutations as described above. Note that in the Py EC F9 mutants, one of the thymine insertions is always located at the 108/220 deletion-duplication junction site of Py EC PCC4 mutants. The region of the polyoma genome comprising nucleotides 62–133 can be folded into a clover-leaf structure (ref. 5 and Fig. 3). The generation of such a structure with circular double-helical DNA is certainly subject to topological constraints. However, transcription activation postulated for prokaryotic replicons¹⁴ or the action of a DNA gyrase (DNA topoisomerase II)^{15,16}, could permit the local melting of the DNA double strand and the formation of a secondary structure with single-stranded DNA. The structure can be stabilized by interaction with specific proteins. Alternatively, a secondary structure on the DNA level could be limited only to stem II (Fig. 3). It is striking that this part of the genome is modified in both Py EC PCC4 and Py EC F9. In Py EC F9, all the mutations are localized between nucleotides 63 and 121 (see Figs 3, 4) and touch mainly stem II of the clover-leaf structure, with the insertion of an adenine facing a complementary thymine. Such an insertion of an extra A·T base pair in a G+C-rich stem could change its interaction with proteins.

Whereas transcription *in vitro* with eukaryotic RNA polymerase II requires mainly the TATA box (D. Hogness and M. Goldberg, unpublished observations), several pieces of experimental evidence suggest that at least 100–200 nucleotides upstream of this sequence are important for transcription regulation *in vivo*^{17,18}. The rearranged region in the polyoma EC mutants is located 100–200 nucleotides before the hetero-

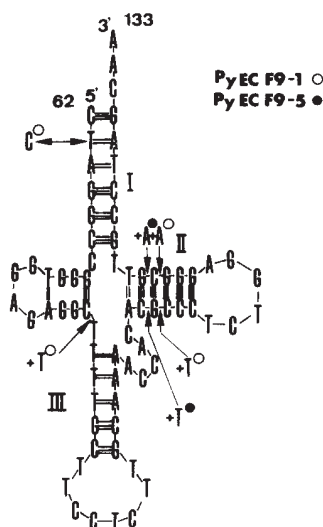


Fig. 3 A clover-leaf structure in the late noncoding region in the polyoma virus genome. The structure suggested by Soeda *et al.*⁵ for wild-type polyoma virus between nucleotides 62 and 133 on the late strand of the genome is drawn. It is possible to draw the complementary structure with the early strand. The insertions and the transition in this region in the two mutants analysed are shown. Note that most insertions for both mutants occur on stem II, always with a double insertion of the complementary adenine and thymine. Note also that in all the Py EC PCC4 mutants, the junction between the deletion and the duplication at 108/220 occurs again on the stem II¹².

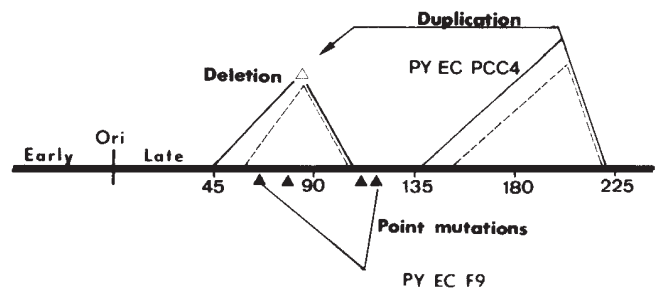


Fig. 4 Summary of the genomic changes around the origin of DNA replication in Py EC PCC4 and Py EC F9 mutants. The upper part of the diagram shows the rearrangements in the Py EC PCC4 mutants where a region between nucleotides 46 or 77 to 107 is deleted and replaced by the region between nucleotides 138 or 157 to 220 which is duplicated (see also ref. 12). The lower part concerns the Py EC F9 mutants with the point mutations (▲) between nucleotides 63 and 121 as described in the text.

geneous 5' ends of the early mRNA (R. Kamen, personal communication). In addition, this same region precedes the 5' ends of late mRNA and should be important for the control of late transcription. Thus, the sequence rearrangements that can occur without impairing the viability of the virus are limited. Previous studies have shown that wild-type polyoma DNA penetrates the EC cells but does not express early mRNA. Recent results¹⁹ reveal that on polyoma EC viral infection of PCC4 of F9, no new sites for early polyoma transcription initiation could be detected but the overall rate of specific polyoma transcription is greater than that of wild-type polyoma in these cells.

A model could be proposed in which a regulatory protein present in EC cells but not in differentiated derivatives inhibits early transcription of the virus by binding to the nucleotide 63–132 region. The modification of the sequence either by point mutations or by rearrangement destroys the specific binding site but still conserves a stable secondary structure otherwise needed for viral expression, thus relieving the inhibition of transcription. Alternatively, another model could be constructed based on a positive regulation. Interestingly, the putative tRNA-like structure in which the mutations occur is included in the region devoid of nucleosomes (gap) in viral chromatin (P. Herbolme *et al.*, unpublished results).

We thank C. Kress for assistance, C. Maczuka for help in the preparation of the manuscript, and A. Minty for valuable criticism of this manuscript. This work has been supported by grants from the French Centre National de la Recherche Scientifique, the Institut National de la Santé et de la Recherche Médicale and the Délégation à la Recherche Scientifique et Technique.

Received 11 November 1980; accepted 6 February 1981.

- Swartzendruber, D. E. & Lehman, J. M. *J. cell. Physiol.* **85**, 179–187 (1975).
- Boccaro, M. & Kelly, F. *Annls Microbiol. Inst. Pasteur, Paris* **A129**, 227–238 (1978).
- Jakob, H., Boon, T., Gaillard, J., Nicolas, J. F. & Jacob, F. *Annls Microbiol. Inst. Pasteur, Paris* **B124**, 269–282 (1973).
- Shermann, M. L. in *Teratomas and Differentiation* (eds Sherman, M. C. & Solter, D.) 189–205 (Academic, New York, 1975).
- Soeda, E., Arrand, J. R., Smolar, N., Walsh, J. E. & Griffin, B. E. *Nature* **283**, 445–453 (1980).
- Seif, R. & Cuzin, F. *J. Virol.* **24**, 721–728 (1977).
- Strickland, S. & Mahdavi, V. *Cell* **15**, 393–403 (1978).
- Pope, J. M. & Rowe, W. P. *J. exp. Med.* **120**, 121–128 (1964).
- Vasseur, M., Kress, C., Montreau, N. & Blangy, D. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1068–1072 (1980).
- Arrand, J. R. *et al. J. Virol.* **33**, 606–618 (1980).
- Magnusson, G. & Nilson, M. G. *J. Virol.* **22**, 646–653 (1977).
- Katinka, M., Yaniv, M., Vasseur, M. & Blangy, D. *Cell* **20**, 393–399 (1980).
- Segal, S., Levine, A. J. & Khoury, G. *Nature* **280**, 335–337 (1979).
- Inokuchi, H., Dove, W. F. & Freifelder, D. *J. molec. Biol.* **74**, 721–727 (1973).
- Liu, L. F., Liu, C. C. & Alberts, B. M. *Cell* **19**, 697–707 (1980).
- Hsieh, T. & Brutlag, D. *Cell* **21**, 115–125 (1980).
- Grosschedl, R. & Birnstiel, M. L. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1432–1436 (1980).
- Benoist, C. & Chambon, P. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3865–3869 (1980).
- Dandolo, L. *et al.* (in preparation).
- Berkner, K. L. & Folk, W. R. *J. biol. Chem.* **252**, 3176–3184 (1977).
- Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499–560.
- Sanger, F. & Coulson, A. *FEBS Lett.* **87**, 107–110 (1978).
- Laskey, R. A. & Mills, A. D. *FEBS Lett.* **82**, 314–318 (1977).